VOLUME 18, 1956/57, No. 4/5 pp. 101-144 Published 20th October 1956

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

A MAGNETODYNAMIC GRAMOPHONE PICK-UP

I. CONSTRUCTION.

by N. WITTENBERG. 681.84.081.48:621.395.623

Professional gramophones are commonly equipped with pick-ups of the electrodynamic type (moving coil and fixed magnet). A magnetodynamic type of pick-up will shortly be marketed in which the coils remain fixed while the magnet moves. The new pick-up has various advantages compared with the electrodynamic type, greater sensitivity and lower price amongst them, which make the new pick-up suitable for non-professional equipment. The article below deals the construction of the magnetodynamic pick-up. Its frequency characteristic and some other properties will be treated in a further article.

Of the various types of gramophone pick-up in use the most common is the piezo-electric type [1]). For professional purposes (in broadcasting studios, and for testing purposes in the manufacture of gramophone records) use is often made of the electrodynamic system: in this, a small coil located between the poles of a permanent magnet follows the movement of the gramophone needle, the signal voltage being induced in it as a result. The construction of the system is in principle the same as that of a moving-coil meter, the gramophone needle taking the place of the pointer moving across the scale.

There are disadvantages inherent in the electrodynamic system. They are manifested particularly when it is desired to employ the system in pick-ups for non-professional equipment of which a high fidelity is required. In such cases the coil must be of extremely light construction, in order to keep the mass of the moving system of the pick-up down to a minimum; otherwise, the resulting accelerations (which may be hundreds of times greater than the acceleration g due to gravity [2])) will give rise to inertia forces that will cause severe wear of the gramophone record. To keep wear within acceptable bounds, the mass of the moving system must not be more than a few milligrams. This implies a coil of extremely thin wire with only a few turns. Consequently the voltage induced in the pick-up is only a fraction of a millivolt. In order to get a good signal-to-noise ratio at the output of the amplifier in spite of this, the amplifier must be provided with a step-up input transformer which has been carefully screened against external magnetic fields (e.g. from the gramophone motor or the amplifier mains transformer). This all adds to the cost.

In order to increase the voltage induced in the moving coil the magnetic field in which the coil moves should be as strong as possible: a rather large permanent magnet is therefore required. The magnet is necessarily situated immediately above the record and hence above the turntable on which the record lies. The turntables of non-professional gramophones are frequently of iron, this being a cheap material and at the same time a heavy one, giving a flywheel action. The magnet exercises an attraction on the iron turntable, increasing the needle pressure on the disc. (The weight of the pick-up may be balanced by a spring or counterpoise.) Often the turntables of non-professional gramophones have a diameter of less than 12″; during the playing of the outer part of a 12″ disc, the pick-up is not directly over the turntable and the needle pressure is lower than it is further inwards, where the pick-up is directly above. The variation in the vertical needle

[1]) See for example L. Alons, New developments in the gramophone world, Philips tech. Rev. 13, 134-144, 1951/52.
[2]) J. L. Ooms, Philips tech. Rev. 17, 101-109, 1955/56.

force sometimes amounts to several grams in ordinary electrodynamic pick-ups. The increase in the force exerted by the needle results in heavier wear of both needle and disc, and causes losses in the reproduction of high frequencies.

The disadvantages of the electrodynamic system — the need for an input transformer and the unnecessarily high needle pressure on part of the gramophone record — are entirely absent in the magnetodynamic pick-up, which will now be described.

Design of the magnetodynamic pick-up

In the magnetodynamic pick-up, the magnet and coil exchange their rôles; the magnet moves while the coil is fixed to the body of the pick-up.

A diagram illustrating the principle is given in *fig. 1*. A rod-shaped permanent magnet *M* is located in the air-gap of a yoke *J* of magnetic material, on which the coils *S* are wound. The rod is magnetized in the direction of the arrow, ***perpendicular to a plane through its axis***, and can turn about that axis, being held in two bushes *P* and *R*. At its lower end the rod has an arm *L* (the needle arm) fixed to it, and this arm carries the needle *N*. The lateral movement of the needle point as it follows the modulated groove in the disc, causes the magnet to turn about its axis.

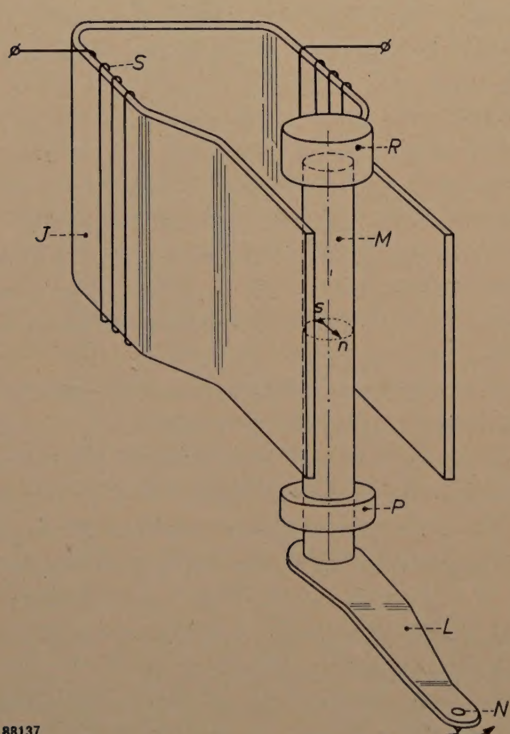


Fig. 1. The essentials of the magnetodynamic pick-up. *M*: ferroxdure rod, magnetised in the direction of the arrow *s-n*. The rod magnet is located between the ends of the yoke *J* carrying the coils *S*, by means of the rubber bush *R* and the polyvinyl chloride bearing *P*, which allow it an angular degree of freedom. The needle arm *L* converts the lateral movement of the needle point *N* into an angular displacement of the magnet about its axis.

The upper bush *R* is made of rubber. The magnet fits into it tightly and thus suffers a restoring couple when displaced that gives it a definite position of equilibrium. The equilibrium position is made to coincide as far as possible with the position of magnetic symmetry (*fig. 2a*), in which the lines of force form a symmetrical pattern, some passing through the air-gap and the ends of the yoke, others through the air only. If the magnet is turned from the equilibrium position the symmetry of fig. 2*a* is upset, and a portion of the flux passes right round the yoke to complete its circuit (fig. 2*b*). The direction of the flux passing round the yoke is dependent on the direction in which the magnet is turned. Lateral movement of the needle point thus produces alternations of flux through the yoke and these induce the signal voltage in the coils.

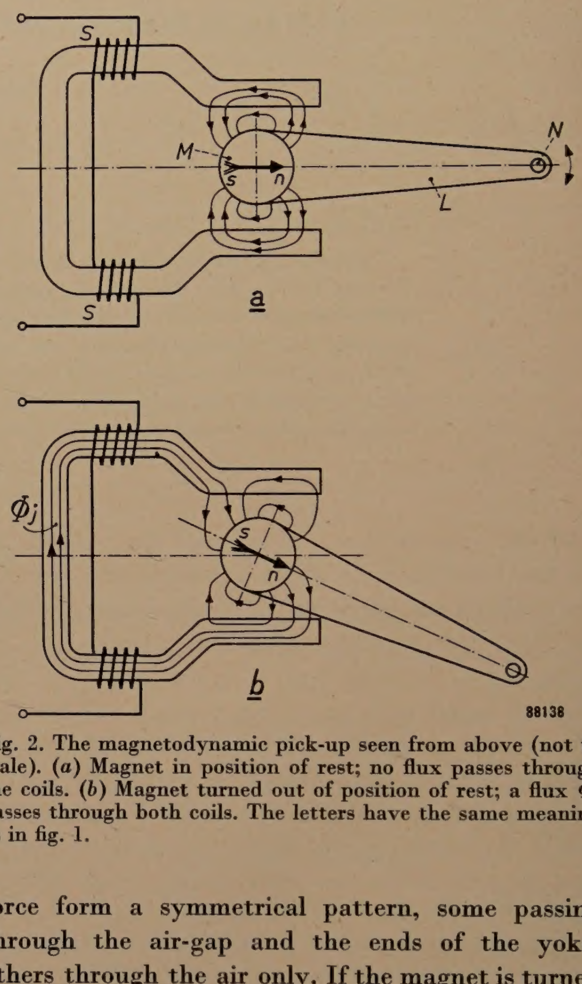


Fig. 2. The magnetodynamic pick-up seen from above (not to scale). (*a*) Magnet in position of rest; no flux passes through the coils. (*b*) Magnet turned out of position of rest; a flux Φ_j passes through both coils. The letters have the same meaning as in fig. 1.

Two important properties of the above arrangement may be deduced from the following experiment. The magnet, driven by a small motor, is made to *rotate* between the ends of the yoke, and the voltage set up in the coils is examined. It is found that the voltage is very nearly sinusoidal, this being due to the relatively large air-gaps. From this fact one may

anticipate that the signal voltage induced by the small angular movement of the magnet during actual use will exhibit only extremely slight deviations from linearity (the magnet departs from its position of symmetry by not more than 1° for standard 78 r.p.m. records, and by not more than 0.5° for microgroove records). In the matter of linearity, then, the magnetodynamic pick-up does not fall short of the electrodynamic type, which in theory is strictly linear.

Of the total flux provided by the magnet, a portion Φ_m reaches the yoke, while the remainder completes its entire circuit in the air as a stray flux. The second property that may be deduced from the experiment with the rotating magnet is the magnitude of the flux Φ_m. For this we require only to know the speed with which the magnet rotates (n in revs. per sec) and to measure the induced voltage E (r.m.s. value). The flux is then given by

$$\Phi_m = \frac{E}{4Fnw} \text{ volt.second,}$$

where F is the shape factor of the voltage function (for a sine wave, $F = \pi/2\sqrt{2} \approx 1.11$) and w is the total number of turns on the coils. For $n = 50$ revolutions per second and $w = 4000$, we find a voltage $E = 0.65$ V, from which it follows that $\Phi_m \approx 0.7$ µVsec. This method can be employed in manufacture for testing the magnets.

From the fact that the rotating magnet produces a sinusoidal voltage it follows that Φ_j, the flux passing through both coils when the magnet is displaced through an angle α from the position of symmetry (*fig. 3*), can be written as $\Phi_j = \Phi_m \sin \alpha$. If y is the displacement of the needle point corresponding to α, and l is the distance between the axes of magnet and needle, $\sin \alpha = y/l$. Hence

$$\Phi_j = \frac{y}{l}\,\Phi_m.$$

If the motion of the needle point is sinusoidal, that is, if $y = \hat{y} \sin \omega t$, then $\Phi_j = (\hat{y}/l)\Phi_m \sin \omega t$. This flux induces an alternating voltage e in the coils given by

$$e = -w\frac{d\Phi_j}{dt} = -\frac{w\Phi_m\hat{y}\omega}{l}\cos\omega t =$$

$$-\frac{w\Phi_m\hat{v}}{l}\cos\omega t, \quad \ldots \ldots \quad (1)$$

where $\hat{v}$ is the peak value of the velocity of the needle point.

The ratio between the voltage and the velocity is termed the sensitivity of the pick-up; it is usual to take the r.m.s. value of the voltage in millivolts and the peak value of velocity in cm/sec. Insertion of the values $w = 4000$, $\Phi_m = 0.7 \times 10^{-6}$ V sec, $l = 0.5$ cm in equation (1) gives us a sensitivity of about 4 mV per cm/sec. For $\hat{v} = 5$ cm/sec — a typical

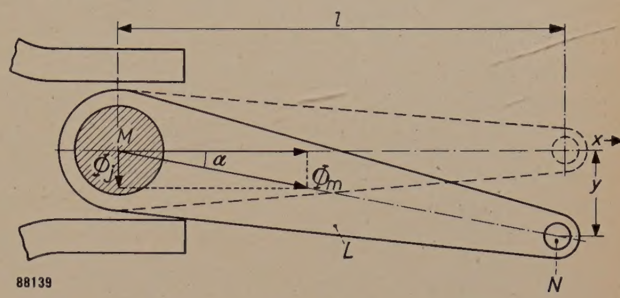


Fig. 3. The same situation as in fig. 2*b* with the main dimensions drawn to scale. Φ_m: the fraction of the total magnetic flux that reaches the yoke. Φ_j: the component of Φ_m that passes through the coils.

value for microgroove records — this sensitivity gives an r.m.s. voltage of 20 mV. The sensitivity is considerably better than that of electrodynamic pick-ups (without input transformer), but not so good as that of the piezo-electric type. Hence, if the magnetodynamic pick-up is to be connected up to a conventional radio receiver, an extra amplification stage is necessary. A suitable pre-amplifier using a transistor will be discussed in a subsequent article.

Construction of the pick-up

The various parts that go to make up the pick-up are shown in *fig. 4*. We shall now examine these components in some detail and consider the requirements they have to satisfy.

The magnet

It follows from equation (1) that, in order to obtain a high voltage from the pick-up (so as to get a high signal-to-noise ratio in the amplifier), the flux Φ_m of the magnet must be made as strong as possible. The magnet must further satisfy the following requirements:

1. Its moment of inertia about the axis should be as small as possible in view of the large accelerations occurring.
2. It should be proof against demagnetization by external magnetic fields. One reason for this is that the magnet together with the needle has to be replaced when the needle is worn out; it must not be possible for the loose magnets to become demagnetized by contact with iron parts or tools or by the action of stray fields from transformers, etc.

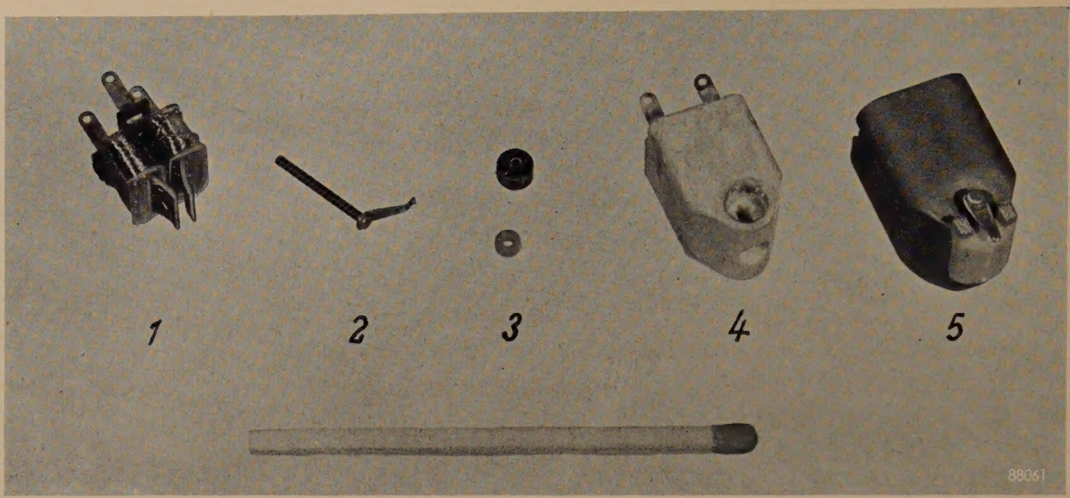


Fig. 4. Component parts of the magnetodynamic pick-up. *1* yoke with coils; *2* magnet with needle arm and needle; *3* rubber bush and p.v.c. bearing; *4* block of polyester resin which is cast around the yoke and coils, place being left for the insertion of magnet and bearings; *5* the block *4* complete with bush, bearing and magnet and shrouded in metal housing for magnetic and electrostatic screening; on either side of the needle arm are protective shoulders.

3. The magnet must be easy and cheap to manufacture.

In order to give the magnet a small moment of inertia an unusual design was adopted: a thin rod magnetized perpendicular to its axis. A "short" magnet such as this has a high demagnetization factor and this design was therefore made feasible only by using a material that has a high coercive force. Ferroxdure [3]) is such a material. Its coercive force is about 80 kA/m (about 1000 oersted), in comparison with about 50 kA/m for "Ticonal" steel [4]). Ferroxdure is suitable for the purpose in mind in other respects as well. Its density is about 4000 kg/m^3, as against about 7000 kg/m^3 for "Ticonal" (a low density is desirable from the point of view of a small moment of inertia). Thin rods of ferroxdure can be manufactured by extrusion, followed by sintering. These processes lend themselves better to mass production than the casting of rods from magnet steel. After sintering, the ferroxdure rods are ground in a centreless-grinder, since they inevitably become somewhat distorted in the course of the sintering operation. Accurately dimensioned rods are thus obtained which are easily interchangeable and have a restoring couple which is satisfactorily reproducible.

A compromise necessarily has to be made between giving the magnet small mass and making it supply a big flux. The best form has been found to be a rod 1 mm in diameter and 12 mm long, its effective length being 8 mm .The flux has the value derived above, namely 0.7 μVsec; the mass is about 40 mg, and the effective mass at the needle point, which determines the magnitude of the inertia forces set up, is only 3 mg. (The effective mass is the equivalent mass considered to be concentrated in the needle point and possessing a moment of inertia with respect to the magnet axis equal to that of the whole moving system.)

The attraction between an iron turntable and a magnet of this small size is quite negligible compared with the minimum force with which the needle must rest on the record in order not to jump the groove. This constitutes one of the great advantages of the magnetodynamic system over the electrodynamic system.

The needle arm and the needle

The function of the needle arm is to communicate to the magnet the movements of the needle point. There are two components of the latter: a lateral movement corresponding to the modulation of the groove, and a vertical movement resulting from the so-called "pinch-effect" (a sinusoidal groove, for example, is narrower in the flanks than at the peaks, with the result that the needle is borne up in the former and sinks down in the latter; see article cited in footnote [1])). Only the *lateral* movement of the needle has to be communicated (without distortion or loss) as an angular movement of the magnet; this implies that the needle arm must be so rigid in a lateral sense that resonance of the needle arm and magnet in the lateral direction occurs only at a frequency higher than the range that has to be

[3]) J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, Philips techn. Rev. **13**, 194-208, 1951/52.
[4]) B. Jonas and H. J. Meerkamp van Embden, Philips tech. Rev. **6**, 8-11, 1941.

reproduced, i.e. above 20 kc/s. At frequencies above resonance almost all the movement of the needle is absorbed by the flexion of the needle arm, resulting in a big drop in the output voltage. In the second part of this article, dealing with the frequency characteristic, it will be shown that the needle arm does in fact satisfy the above condition.

Since the needle extends below the needle arm, the lateral force on the needle point creates a couple that to some extent twists the needle arm. Like the frequency of lateral resonance, the resonant frequency of this torsional effect must lie above the range of frequencies to be reproduced. If the system is excited at the torsional resonance frequency, the magnet remains practically still and the output voltage drops to zero, while the needle makes a shrill sound and there is heavy wear of the record. The frequency at which torsional resonance occurs is decided by an effective mass (other than that for lateral movement) at the needle point and the torsional stiffness of the needle arm. The obvious way of raising the torsional resonance frequency above 20 kc/s is to make the needle arm thick enough. However, this would bring us into conflict with a third requirement that the needle arm has to satisfy: it must not transmit to the magnet the vertical movement of the needle resulting from pinch effect; this vertical movement occurs at a frequency double the modulation frequency and, if it contributed to the pick-up output, would introduce a second harmonic. In theory, of course, no movement of the magnet other than the turning movement induces a voltage in the coils; but the slight departures from symmetry that are inevitable in mass-production prevent this ideal from being completely realised. Hence the magnetodynamic pick-up, like the electrodynamic one, does have a certain response to vertical needle movement.

The "vertical response" can be reduced to a minimum by making the needle arm relatively flexible in the vertical direction, whereby the resonant frequency of magnet and arm for vertical movement is reduced to below the frequencies where pinch effect is considerable (roughly speaking above 1000 c/s); hence the vertical movement is absorbed by the needle arm and thus prevented from affecting the magnet. Flexibility of the needle arm in the vertical direction has two further advantages. Firstly, it means that the effective axis for movement in the vertical plane is closer to the needle, so that the effective mass for movement in this direction is virtually limited to the mass of the needle and its fixture to the arm. Secondly, if by mischange the pick-up should be dropped on to the disc, the needle arm can flex back such that the weight of head is taken on two shoulders fixed to the housing (visible in fig. 4, No. 5); this protects the brittle ferroxdure rod against breakage. On the other hand, however, the needle arm must be sufficiently stiff to support the weight which the pick-up exerts, via the needle, on the record (about 10 grams) without bending too far. By suitable choice of the material, shape and dimensions of the needle arm these requirements can be satisfied. The resulting torsional resonance frequency is in the region of 25 kc/s while the resonant frequency for lateral vibration is still higher.

From what has been said it will be evident that the needle arm, apparently so simple, is a component to which a great deal of attention has to be given.

The magnet and the needle are fixed to the needle arm by cleating them with small collars of aluminium. Needle, collar and needle arm are first positioned in the cleating jig. The unflanged end of the collar is then cleated by a suitably shaped tool (see *fig. 5*) The magnet is fixed to the needle arm by

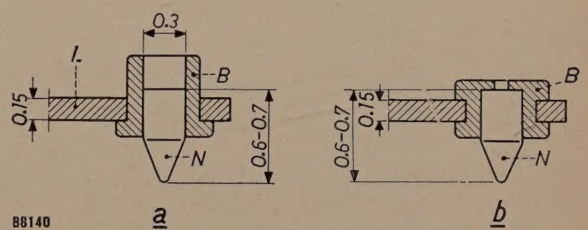


Fig. 5. Attachment of the needle N to the needle arm L by an aluminium cleat B. (a) Before cleating, and (b) after. Dimensions are in mm.

a similar method. Not only does this method give a reliable joint but all play is taken up by the aluminium cleating so that there is no need to demand extremely fine tolerances in the dimensions of the parts.

The magnetodynamic pick-up (*fig. 6*) is manufactured in two models, identified by a coloured mark, which differ only in the type of needle and which are easily interchangeable. Type AG 3020 (with a green spot) is for standard 78 r.p.m. records and has a sapphire needle; the radius of curvature of its tip is 75 microns, as is usual in needles for these records. Type AG 3021 (red spot) is for microgroove records; the tip of its diamond needle has a radius of curvature of 25 microns.

Diamond was chosen for the needle point of the second type for the following reasons. With the force of 10 g weight with which the point rests on the record and the smaller radius of curvature of the microgroove needle, the pressure is so high that a sapphire needle wears comparatively quickly. The wear of the needle results in distorted reproduction, particular-

ly of high notes. If the pick-up is used with an amplifier and loudspeakers capable of reproducing very high audio frequencies [5]), the wearing of the sapphire point is perceptible as a troublesome distortion after comparatively few records have been played. The rate at which diamond wears is some tens of times slower. Added to this, less wear of the needle point means less wear in the records. One might expect that the harder and more resistant to wear the needle point, the greater the wear of the record but in fact the latter is mainly dependent on the shape of the needle point: the harder the needle, therefore, the longer it will retain its initial approximately hemispherical shape and hence the smaller the wear of the record.

Fig. 6. Philips record-changer, type AG 1003, equipped with the AG 3021 magnetodynamic pick-up.

The magnet bushes

The magnet bushes must allow the magnet to turn about its axis, but provision also has to be made for a restoring couple that tends to bring it back to a definite position of equilibrium. This is necessary to ensure that the needle arm lies parallel to the groove when the latter is unmodulated. If this condition is satisfied, tracing distortion is at a minimum [6]).

The restoring couple is obtained by clamping one end of the magnet into a rubber bush (R in fig. 1). The stiffness so obtained must have a certain minimum value for the reason given above, but it must not be greater than necessary; at low frequencies, the stiffness largely determines the force required to give the needle a certain lateral displacement, and this force in turn determines the minimum downward force of the needle required to prevent it jumping the groove. We shall return to this matter of choosing a value for the vertical force later.

A value of 200 N/m ($= 0.2 \times 10^6$ dyne/cm) has been chosen for the stiffness s. The maximum amplitude $\hat{y}_{\max}$ is 100 microns for a standard 78 r.p.m. record at low frequencies; at higher frequencies the maximum amplitude is less, being limited by other factors during the cutting of the record: see article quoted in footnote [2]). The lateral force necessary to give the needle point this maximum amplitude is therefore of the order of $s\hat{y}_{\max} = 200 \times 100 \times 10^{-6}\text{N} = 20$ mN.

Apart from the restoring couple, the moving system must have a definite amount of damping. This is a question not only of the resonance phenomena already mentioned but also of the resonance of the mass of the pick-up arm with the stiffness of the moving system. In Philips record-changers equipped with a magnetodynamic pick-up this latter resonance occurs at about 30 c/s. The damping action of the rubber bush is very slight; for this reason, at the lower end a polyvinyl chloride bush (P in fig. 1) is used in which the magnet can turn. The plastic gives considerable mechanical damping

[5]) See for example J. J. Schurink, The twin-cone moving-coil loudspeaker, Philips tech. Rev. **16**, 241-249, 1954/55.

[6]) To discuss the reasons for this would be outside the scope of the present article. It may be mentioned here that the arm of the pick-up should ideally be designed so that the symmetry plane of the tracking system is always parallel the tangent to the unmodulated groove (see for example to B. B. Bauer, Tracking angle in phonograph pick-ups, Electronics **18**, 110-115, March 1954).

over an extensive range of frequency, but the restoring couple set up by the lower bush is very small. Together with the effective mass and the stiffness, the damping provided by the lower bush goes to make up the mechanical impedance; it must not be made excessive, because a damping force is also supplied by the needle point. This being so, the magnet is not clamped in the polyvinyl chloride bush but can rotate as in a bearing; in fact the bore of the bush is somewhat greater than the diameter of the magnet so that the latter is pulled to one side of the bearing wall by the frictional force between needle and record. This turns out to give a satisfactory degree of damping, the fact that the bush in question is lower down and very close to the needle also playing a part. Interchanging the two bushes has little effect on the restoring couple but causes a big decrease in damping.

The mounting of the magnet described above is not rigid in any direction and hence it allows of movements other than rotation. This is undesirable, as we have already seen, in view of the second harmonic that may arise from the pinch effect. On the other hand, mounting the magnet in bearings that permit a turning movement only appears to make reproduction "hoarse", particularly at high frequencies. There is a rule valid for all pick-up systems according to which the moving system of the pick-up must be given various degrees of freedom in the higher frequency range; the limited movements thereby permitted keep the mechanical impedance to a minimum. As far as rotational movements are concerned, the impedance is a minimum for rotation about that axis for which the moment of inertia is least. At high frequencies the horizontal and vertical movements of the needle point, and hence those of the whole moving system, are very complicated and very difficult to investigate; their amplitudes amount to a few microns at the most and their accelerations attain some hundreds of times g, as already stated. If the moving system is deprived of one degree of freedom, the mechanical impedance at the needle point may rise so high that the forces set up by the groove modulation acquire values at which a distortionless trace is no longer possible. In every pick-up, therefore, a favourable compromise must be sought between the number of degrees of freedom allowed to the moving system and the acceptance of voltages produced by undesired movements of that system. The quality of the pick-up depends considerably on how well this compromise is chosen. In this respect a system such as the magnetodynamic, in which in principle only one type of movement induces a voltage, has a great advantage over a system wherein voltages are induced by more than one type of movement.

The yoke

The yoke must have as small a reluctance as possible, and is therefore made of a material of high initial permeability. It consists of a strip of cross-section of $6 \times 0.35 = 2.1$ mm², bent into the required shape.

If the needle point has an amplitude $\hat{y}$ of 0.1 mm, the amplitude of the alternating flux through the yoke is:

$$\Phi_\mathrm{j} = \Phi_\mathrm{m} \frac{\hat{y}}{l} = 0.7 \times 10^{-6} \times \frac{0.1}{5} = 14 \times 10^{-9} \text{ volt.sec.}$$

Hence the magnetic induction in the yoke in the same circumstances is:

$$B_\mathrm{j} = \frac{14 \times 10^{-9}}{2.1 \times 10^{-6}} \approx 7 \times 10^{-3} \text{ volt.sec/m}^2 \ (= 70 \text{ gauss}).$$

This is a very low figure, and hence the possibility of distortion due to the saturation of the yoke is ruled out (the saturation induction for the material used is about 0.5 Vsec/m²). Equally, there is little likelihood that magnetic asymmetry will produce too high an induction in the yoke, for even if the entire flux Φ passed through it, the static induction in the yoke would still only amount to 0.35 Vsec/m².

The low value of the alternating induction has an advantage over and above the elimination of distortion due to saturation. There is no danger of *noise* arising from the reorientation of Weiss domains (Barkhausen effect [7])). The absence of this kind of noise has been confirmed by an experiment in which the needle was made to move with a large amplitude at a low frequency; there was absolutely no indication of any noise associated with this movement.

Polyester resin is cast around the yoke carrying the coils. A sturdy block (fig. 4, No. 4) is thus formed, in which the positions for the bearings are accurately centred with respect to the yoke. The coils are completely enclosed, making the pick-up proof against tropical climates.

The two coils together constitute an approximately astatic system and hence they are but little affected by stray magnetic fields arising from the gramophone motor and any transformers that may be in the vicinity. On account of the relatively high output voltage, the leads from the pick-up to the amplifier pick up little interference. In both these respects the magnetodynamic pick-up compares

[7]) H. Barkhausen, Phys. Z. **20**, 401-403, 1919; B. van der Pol, Versl. Kon. Akad. Wet. Amsterdam, **29-I**, 341-348, 1920.

favourably with electrodynamic and other magnetic types. Nevertheless, as a further measure against any residual fields, it is provided with magnetic screening. This consists of a soft iron casing (fig. 4, No. 5) into which the polyester block is fixed once the two bearings have been inserted in it.

Finally, the system thus screened is mounted in a "Philite" housing. The housing has a terminal by which the magnetic screening can be connected to the earth terminal of the amplifier, so that it serves as electrostatic screening at the same time. The yoke is also connected to earth, thus obviating undesirable phenomena due to static electricity generated by the friction between needle and record.

The optimum value for the vertical needle force

F_v, the vertical force with which the needle presses on the record, is an important quantity. In general, it is desirable to keep this force small to prevent excessive wear. If the pick-up is too light, on the other hand, there is a risk that the needle will jump the groove. The following considerations show us in what circumstances this risk is present.

Apart from the vertical force F_v, there is a lateral force F_l acting on the needle point, which is the resultant of several lateral forces. Diagrams illustrating two cases appear in *fig. 7*, (*a*) where $F_v > F_l$ and (*b*) where $F_v < F_l$. F_r, the resultant of F_v and F_l, can be resolved into two components, F_{n1} and F_{n2}, perpendicular to the walls of the groove, which are approximately at right angles to each other. In both cases F_{n1} is balanced by a reaction from the left-hand wall on the needle point at Q_1, their point of contact. In fig. 7*a* the direction of F_{n2} is towards the other wall and it is balanced by the reaction at point Q_2, the needle thus being kept in contact with the wall at this point. However, in fig. 7*b* the direction of F_{n2} is away from Q_2; the needle will therefore tend to lose contact with the right-hand wall and, if F_{n2} is sufficient to overcome the frictional force due to movement against the left-hand wall, the needle will ride up the latter. This results in distortion or, at worst, de-tracking of the needle. Hence the highest value of F_l that occurs determines the lowest permissible value of F_v if this danger be avoided.

As already stated, F_l is the resultant of a number of lateral forces. Of these, the inertial force of the moving system (its magnitude is the product of the effective mass at the needle point and its acceleration) is the largest at high frequencies. In order to minimise this force the moving system is designed to have as small an effective mass as possible; the value of the latter is in fact 3 mg.

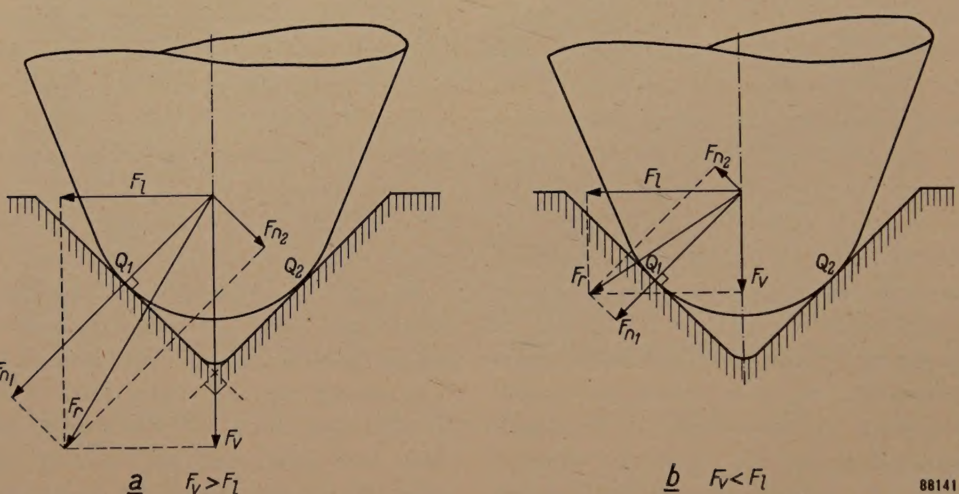


Fig. 7. Forces on the needle point. F_v vertical force with which the needle rests on the disc. F_l total lateral force. In case (*a*) $F_v > F_l$, in case (*b*) $F_v < F_l$. F_r, the resultant of F_v and F_l, has components F_{n1} and F_{n2} perpendicular to the two walls of the groove. In case (*a*) F_{n2} is in a direction such that the needle maintains contact with the right-hand wall at point Q_2; in case (*b*) the direction of F_{n2} is such that this contact is not maintained and the needle rides up along the left-hand wall of the groove. (Friction, and the deformation of the walls under needle pressure, are neglected.)

The maximum permissible acceleration is found by considering the smallest radius of curvature that may occur in the groove without distortion arising; if the radius of curvature ϱ of the modulation peaks were smaller than R, the radius of curvature of the needle point, the needle would be in contact with the groove at three points instead of two, and distortion known as over-modulation would arise. Taking $\varrho = R$ as a rough basis, a peak acceleration

of the order of 300 g is found for the fundamentals of the higher frequencies (see article cited in footnote [2])). However a third harmonic of about 40% is then present in the velocity [8]), and consequently one of $3 \times 40 = 120\%$ in the acceleration. Taking this into consideration, we arrive at a maximum value of the total acceleration of $(100 + 120)\%$ of 300 g, that is, of 660 $g \approx 6600$ m/sec^2. This, with the effective mass of 3 mg, produces a maximum lateral force of inertia of $3 \times 10^{-6} \times 6600$ N ≈ 20 mN (≈ 2 g weight).

Oscillograph measurements of the voltage delivered by the pick-up shows that accelerations even greater than 6600 m/sec^2 — and hence correspondingly greater forces of inertia — sometimes occur at the needle point. The higher accelerations may be caused by higher modulation velocities (where $\varrho < R$), giving overmodulation, which brings about a big increase in distortion; this in its turn again increases the acceleration owing to the harmonics involved. A further cause of the increased accelerations may be the building up of high amplitude needle vibrations due to groove-needle resonance [9]).

There are two other lateral forces besides the force of inertia just stated. Firstly, owing to the frictional drag of the needle on the record and the geometry of the pick-up arm, there is a constant inward force on the pick-up across the record; in Philips gramophones this is about 15 mN. Secondly there is the stiffness force which was shown earlier to have a maximum value of 20 mN at the lower frequencies for normal 78 r.p.m. records, this being the product of the stiffness $s = 200$ N/m and maximum needle displacement $\hat{y}_{max} = 100$ microns. Such large displacements occur only at low frequencies, but the higher frequencies *and* their harmonics can also be present in the groove at the same time. These three forces can therefore be additive and in this way we arrive at a maximum of $20 + 15 + 20 = 55$ mN for the lateral force F_l; this, then, is the lowest value permissible for F_v. There is still a further point to be considered, however: the pinch effect causes a periodic variation in the magnitude of F_v. The amplitude of the alternating component may amount to about 10 mN so that a *static* value of at least $55 + 10 = 65$ mN is required for F_v.

This calculation is by its very nature only an approximation to the real state of affairs. However, bearing in mind that all the unfavourable circumstances are rarely present in combination, the minimum value of 65 mN, as calculated, agrees well with the value found by experiment. In fact a force of from 60 to 70 mN appears to be just sufficient for records with large modulation amplitudes to be faultlessly played, provided that the greatest attention is paid to the bearings and the balancing of the pick-up arm. Extreme care in production is not an insuperable objection in the making of professional equipment but, in the case of gramophones for domestic use, it is required that efficient tracing of the groove be obtained with pick-up arm that is simple and inexpensive. For these reasons F_v is given a somewhat higher value, viz. 100 mN (about 10 g weight). This value not only guarantees stable operation but also minimum wear, for it has been shown experimentally that with a weaker force there is again an increase in wear.

Part II of this article will deal with some properties of the magnetodynamic pick-up, including its frequency characteristic.

Summary. The new magnetodynamic pick-up type AG 3020/21 has a small rod-shaped magnet as its moving system; the rod is magnetized perpendicularly to its axis, about which it can turn, and is mounted between the ends of a yoke of magnetically soft material. A needle arm is fixed to the rod magnet whereby the lateral movement of the needle as it follows the groove in the record is converted into an angular movement of the magnet. An alternating flux is thus produced in the yoke, giving rise to a signal voltage in the coils wound on it. The angular movement of the magnet is provided for by an upper flexible bush of rubber and a bearing at the lower end of polyvinyl chloride. The rubber bush gives the magnet a restoring couple and the p.v.c. bearing provides the necessary damping against undesired resonances. The pick-up is manufactured in two models, type AG 3020 with a sapphire needle fitting the groove in standard 78 r.p.m. discs, and type AG 3021 with a diamond needle for microgroove discs. The departure from linearity between the angular movement of the magnet and the induced signal voltage is extremely slight. The sensitivity of the pick-up (ratio of r.m.s. voltage to peak needle velocity) is about 4 mV per cm/sec. A study of F_v, the force with which the needle presses on the disc, yields an optimum value of F_v of 100 mN ($\approx$ 10 g weight) in gramophones for domestic use, this value being sufficient to prevent de-tracking and giving minimum wear. The pick-up is proof against tropical climatic conditions.

[8]) J. A. Pierce and F. V. Hunt, On distortion of sound reproduction from phonograph records, J. Acoust. Soc. Amer. **10**, 14-28, 1938/39; W. D. Lewis and F. V. Hunt, A theory of tracing in sound reproduction from phonograph records, J. Acoust. Soc. Amer. **12**, 348-365, 1940/41.

[9]) J. B. S. M. Kerstens, Mechanical phenomena in high-note reproduction by gramophone pick-ups, Philips tech. Rev. **18**, 89-97, 1956/57, in particular pp. 000. See also Part II of present article.

THE NOISE EMISSION OF BALLASTS FOR FLUORESCENT LAMPS

by E. W. VAN HEUVEN.

621.327.534.15.032.434:
621.318.424:534.833

The introduction of fluorescent lighting into offices, shops and the home, required a reduction of the noise emission to a level compatible with the "silence" desirable in such surroundings. This article describes how the noise problem was solved, by special methods of construction and mounting and by acoustic insulation. Both the theoretical basis of these solutions and the associated measurements are discussed.

Unlike incandescent lamps, gas-discharge lamps, such as tubular fluorescent lamps ("TL" lamps), cannot be directly connected to the mains but must be put in series with some type of ballast to limit the current in the discharge tube. [1])

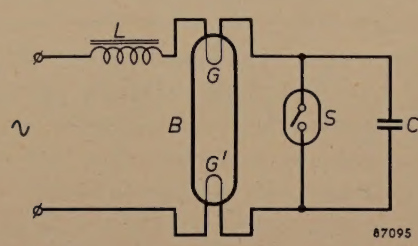


Fig. 1. Diagram of a fluorescent lamp with ballast. *B* lamp with filaments *G*, *G'*, *L* choke, *S* starter with interference-suppression capacitor *C*.

A frequently-used circuit, in which the ballast consists of a self-inductance (choke), is shown diagrammatically in *fig. 1*. When the lamp is switched on, the two filaments *G* and *G'* of the discharge tube are connected in series with each other and with the choke via the starter *S*. The discharge circuit is then short-circuited and a fairly large current passes through the filaments, which quickly heat up. When the connection is broken at *S* (this happens automatically), the discharge is initiated. After each half phase of the alternating current the lamp extinguishes and is re-ignited, now, however, without the intervention of the starter since, on account of their thermal inertia, the filaments are still sufficiently hot, and moreover, there are sufficient residual ions and electrons in the gas to initiate the discharge at a voltage only slightly higher than the working voltage.

With the increasing use of fluorescent lamps for indoor lighting (offices, living-rooms) more attention came to be paid to the noise associated with these installations. The noise, which includes both low notes (hum) and high notes (rustle), originates principally in the ballasts. The cause lies in the alternating magnetization and the consequent deformation of the iron core of the ballast.

We shall first discuss the nature of the noise and the manner in which vibrations of the ballast are communicated to the air. We shall then go on to deal with the theoretical aspects of core deformations, confining ourselves for the time being to low frequencies. From these considerations, the precautions that can be taken to combat ballast hum are derived. Lastly, we shall deal more fully with the causes of rustle and how it can be combated. [2])

The nature of the noise

Fig. 2 shows oscillograms recorded on a fluorescent lamp with a choke; *I* is the voltage across the lamp, *II* the voltage across the choke, *III* the current flowing through the lamp and consequently through the choke. To a first approximation, this current is sinusoidal. Closer examination shows that there are certain deviations from the sinusoidal form, so that in addition to the mains frequency (50 c/s), higher frequencies occur, which are odd multiples of the mains frequency.

The changes in the shape of the core of the coil are, as we shall see below, periodic, having a fundamental frequency twice that of the mains (i.e. 100 c/s). Apart from this frequency, "higher harmonics" occur in the vibrations of the core, whose frequencies are small multiples of 100 c/s. These low-frequency vibrations are responsible for *hum*, a noise that is widely encountered with alternating current apparatus (it can be heard, for example, if the ear be placed close to a transformer housing).

[1]) For a description of ballasts, see Fluorescent Lighting, edited by C. Zwikker, Philips Technical Library, 1952.

[2]) See also: E. W. Heuven, On the noise of fluorescent lighting installations, Acustica 5, 101-111, 1955 (no. 2).

Superimposed upon the approximately sinusoidal current there are other irregularities which take the form of small current peaks of short duration. This may be seen from the rapid voltage variations found

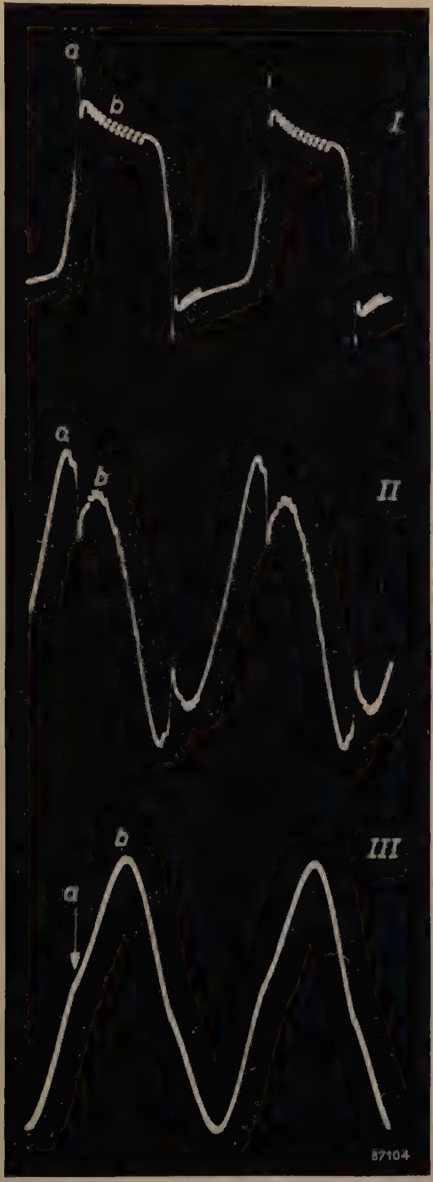


Fig. 2. Oscillograms of the voltage across the lamp (*I*), the voltage across the choke (*II*) and the current through the lamp and the choke (*III*). Re-ignition occurs at *a*, giving rise to voltage peaks. The voltage peaks at *b* result from irregularities in the discharge. Both phenomena (*a* and *b*) give rise to variations in the current (barely discernible in *III*).

in curves *I* and *II* in fig. 2 in two places, viz. *a*) at the moment of re-ignition, *b*) during the time that the current is large. (The corresponding current peaks are not sufficiently large to be visible in the oscillogram of the current, curve *III*, owing to the high impedance of the choke for rapid variations.) Peaks (*b*) can have two causes, viz. irregular discharges in the gas around the filaments, and plasma oscillations. Current peaks (*a*) and (*b*) are responsible, as will be described in greater detail later, for higher frequency vibrations in the core (1000-3000 c/s), which give rise to the noise known as *rustle.*

In addition to the noises designated hum and rustle, a third troublesome noise sometimes occurs, which might be called "rattle". This is likewise caused by mechanical vibration of the ballast and its fixture, when these contain mutually supporting parts that are held together only by gravity. If, as a result of vibration, vertical accelerations occur somewhere in the construction that are greater than "g" the acceleration due to gravity (for a frequency of 100 c/s, this occurs when the amplitude is greater than 25 microns), these parts move with respect to each other and cause rattle. Since this noise is less important, and can usually be easily rectified, we shall not discuss it further.

Transfer of vibrations from the ballast to the air

The way in which the vibrations of the ballast reach the ear is different for hum than for rustle. Sound frequencies of the order of 100 c/s correspond to wavelengths of several metres, i.e. much greater than the dimensions of the ballast. A ballast suspended freely in space (or elastically suspended) would therefore be a very poor radiator of such waves [3]), and would give very little hum. It is more customary, however, to attach the ballast firmly to a wall, to the ceiling or to a fixture. The surfaces of the latter objects, which are brought into sympathetic vibration with the ballast, are not small compared with the wavelengths mentioned, so that they will emit a perceptible hum. The same phenomenon is involved here as in the well-known experiment with a tuning fork. The fork vibrating freely in the air is scarcely audible, but is clearly audible if its stem is held on a table top or similar surface.

With rustle the frequencies involved are of the order of 3000 c/s, corresponding to wavelengths of about 10 cm; the ballast itself is a suitable radiator for such notes, so that rustle will be audible even in the absence of any sounding-board.

To avoid noise from the ballast it is therefore necessary to prevent vibrations from being transmitted to walls, etc. The hum is then inaudible. To combat rustle it is necessary, in addition, to suppress direct emission by the ballast in the frequency range 1000-3000 c/s. We shall now analyze the phenomenon of hum in greater detail.

The origin of hum

The deformation of an iron core by magnetization may result from:

[3]) See, for example, A. Th. van Urk and R. Vermeulen, The radiation of sound, Philips tech. Rev. 4, 213-222, 1939.

a) elastic deformation in iron cores with an (air) gap due to attraction between the parts of the iron circuit on either side of the gap.

b) magnetostriction.

The term magnetostriction comprises a whole series of deformation phenomena, shown by ferromagnetic materials on magnetization (as well as, in a number of cases, the reciprocal effects, i.e. change of the magnetic properties by the application of a mechanical load). The most important of these effects is the change in length in the direction of magnetization [4]). The relative change in length $\varepsilon = \Delta l/l$ is positive for iron at low magnetizations, but changes in sign as the magnetization approaches the saturation value. *Fig. 3* gives a general idea of ε as function of the magnetization J or, what amounts to practically the same thing, as a function of the induction B [5]).

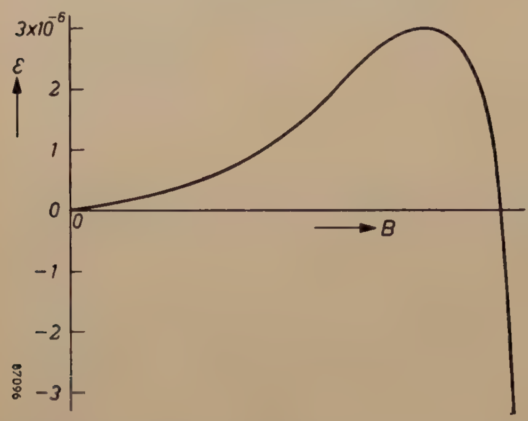


Fig. 3. Magnetostriction of transformer lamination. $\varepsilon = \Delta l/l$ is plotted as a function of the induction B. As saturation is approached ε changes sign.

The beginning of the curve is approximately described by the formula:

$$\varepsilon = \frac{a}{2\mu_0} B^2, \quad \ldots \ldots \quad (1)$$

where $\mu_0 = 4\pi/10^7$ volt seconds/ampere metre, and a is a constant, which for transformer iron, depending upon the silicon content and the pre-treatment of the iron, amounts to $1\text{-}3 \times 10^{-12}$ metres per newton (m/N). The induction in a choke core is seldom very high in practice, so that formula (1) is usually valid. It may be seen from the formula that the magnetostrictive change in length is approximately sinusoidal and has a frequency double that of the current, which is considered to be sinusoidal. The induction B then varies [6]) as $\sin^2\omega t$, i.e. as $\frac{1}{2}(1 - \cos 2\omega t)$.

If care is taken in designing the iron core to ensure uniform magnetization, then the magnetostictive expansion leads to an approximately proportional increase in the external dimensions in a plane parallel to that of the laminations [7]). Since at the corners of the core the induction is somewhat greater on the inside of the bend than on the outside, there occurs at the same time a slight bending, which we shall ignore for the time being.

Measures to reduce hum

Having regard to the contribution of elastic deformation (*a*), we must distinguish between iron cores with air-gaps (for chokes) and cores without air-gaps (for transformers). We shall first examine the simplest case: symmetrical cores with air-gaps.

Symmetrical cores with air-gaps

Consider a core of the design shown in *fig. 4*. The air-gap is located in a plane of symmetry of the core which passes through the centre of gravity. We shall assume for the moment that the ballast is freely suspended in space or is elastically suspended in such a way that it can move freely, the centre of gravity thus remaining stationary.

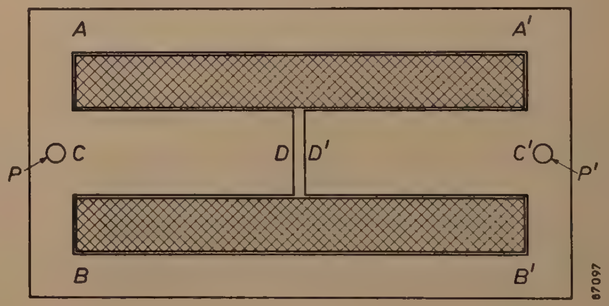


Fig. 4. Cross-section of a choke with a symmetrical iron core. The core consists of three limbs AA', BB' and CC' connected by yokes. The air-gap is between D and D' in the centre limb and can be filled with non-magnetic material. P and P' are the points of attachment.

---

4) W. D. Williams, Magnetostrictive phenomena, General Electric Rev. **45**, 161-163, 1942. S. C. Leonard, Magnetostriction made visible, idem **45**, 637-641, 1942.

5) For other ferromagnetic materials, or for single crystals of iron oriented in a particular way, these curves may be different: see the literature cited in 4).

6) If the (odd) higher harmonics are taken into account in the expression for the current ($i = A \sin \omega t + B \sin 3\omega t + C \sin 5\omega t + \ldots$), then on squaring such terms as $\sin 4\omega t$ are found, i.e. frequencies that are multiples of double the mains frequency.

7) Together with the elongation by a factor $(1 + \varepsilon)$ in the direction of magnetization, there occurs a contraction by a factor $(1 - \varepsilon/2)$ at right angles to the direction of magnetization (magnetostriction causes hardly any variation in volume). The transverse contraction of the yokes of a core, see fig. 4, thus acts in opposition to the lengthwise expansion of the limbs, and vice versa. Nevertheless positive magnetostriction results in an increase in the periphery of the laminations since the limbs and yokes are appreciably longer than they are broad.

When the iron is magnetized, the poles on either side of the gap attract each other and the two halves CD and $C'D'$ of the centre limb are stretched (the other sections of the iron circuit are also deformed: the outer limbs AA' and BB' are compressed and the yokes AB and $A'B'$ are bent). However, the elastic deformation is counteracted by the magnetostriction, for the latter tends to increase the dimensions of the core in the plane of the laminations (plane of the diagram in fig. 4), that is to say, magnetostrictive deformation implies that points on the line CC' move away from the centre of gravity, the displacement being proportional to the distance to the centre of the gap. With elastic deformation it is just the reverse: here the displacement is greatest at the gap and less towards the outside. On line CC' therefore there should be two points — located symmetrically with respect to the centre of gravity — where the total displacement is nil. It has been found that the elastic and the magnetostrictive deformations depend in an identical manner upon the induction (and, therefore, upon the current through the coil), so that during the whole vibration, resulting from variations of the current, both points mentioned remain stationary. In practice the elastic displacement is found to be greater than the magnetostrictive displacement at all points within the core, so that these nodal points lie outside the core.

If the air-gap now be filled with a material with the same magnetic properties as air, but whose rigidity is such that the elastic deformation is reduced, then the vibration amplitudes of all points in the core are reduced, giving a considerable reduction in the hum. Moreover, the two nodal points are brought closer to the centre of gravity, and may even lie within the core. The location of these points is governed by the dimensions and by the filling material which can be so chosen that these points exactly coincide with P and P', the points of attachment of the choke. Vibration cannot then be transmitted to the wall.

The above ideas, when worked out quantitatively, lead to the following formula for the increase in distance between two arbitrary points P and P' situated symmetrically at a distance L from each other on the centre limb:

$$\Delta L = \frac{B^2}{2\mu_0}\left[\alpha L - \left(\frac{O_1}{S_1+S_2}\right)\left(1-\frac{S_1}{O_1E_1/L}\right)\right]. \qquad (2)$$

where O_1 is the cross-section, E_1 Young's modulus of the material of the core, and S_1 and S_2 are the respective "rigidities" of the core, and the gap-filling material. In this formula $S_2 = O_2E_2/\delta$ should be regarded as the unknown (O_2 = cross-section and E_2 = Young's modulus of the filling material, δ = the gap width). S_2 must be given such a value, by a suitable choice of material, that $\Delta L = 0$.

In deriving (2) it has been assumed that the elastic and magnetostrictive deformations may be calculated as in the static case. For low hum frequencies this is justified (the forces and the deformations are in phase).

The poles on either side of the air-gap (at D and D') are drawn together by a force

$$F = \tfrac{1}{2}BHO_1 = \frac{1}{2\mu_0}B^2O_1 \quad \ldots\ldots \quad (3)$$

(H is the magnetic field in the gap).

By filling the gap with an elastic material, D and D' are pushed away from each other with a force

$$F_2 = -\frac{O_2E_2}{\delta}\Delta\delta = -S_2\Delta\delta. \quad \ldots\ldots \quad (4)$$

Here, $\Delta\delta$ is the increase in the distance δ between D and D'.

The resultant force $F - F_2 = F_1$ gives rise to a complex deformation of the core: limbs AA' and BB' are compressed and bent, parts AB and $A'B'$ are mainly bent and parts CD and $C'D'$ are elongated. These deformations result in a reduction $-\Delta\delta$ of δ at the level of the gap. $-\Delta\delta$ is proportional to F_1:

$$F_1 = -S_1\Delta\delta. \quad \ldots\ldots\ldots \quad (5)$$

By analogy with equation (4), the proportionality factor S_1 can be regarded as a rigidity. Owing to the complex character of the deformation corresponding to S_1, the latter cannot be expressed simply in other quantities. Combination of (3), (4) and (5) gives:

$$\Delta\delta = -\frac{B^2}{2\mu_0}\left(\frac{O_1}{S_1+S_2}\right). \quad \ldots\ldots \quad (6)$$

The extension of each of the parts CD and $C'D'$ is such, that two points in such a part, originally separated by a distance l, are at a distance $l + \Delta l$ from each other, where

$$\Delta l = l\frac{F_1}{O_1E_1}.$$

If we assume that points P and P' (separated by a distance L, of which $L - \delta$ is in iron and δ in the filling material) can still be regarded as belonging to the centre limb, then for the increase in L as a result of the elastic deformation, we have:

$$\Delta L_{\text{elast}} = \Delta\delta + \frac{L-\delta}{O_1E_1}F_1, \quad \ldots\ldots \quad (7)$$

or, approximately, since $\delta \ll L$:

$$\Delta L_{\text{elast}} = \Delta\delta\left(1-\frac{L}{O_1E_1}S_1\right) = -\frac{B^2}{2\mu_0}\left(\frac{O_1}{S_1+S_2}\right)\left(1-\frac{L}{O_1E_1}S_1\right). \quad (8)$$

Superimposed upon the elastic deformation is the magnetostrictive deformation, which according to equation (1) is

$$\Delta L_{\text{magn}} = \varepsilon L = \frac{\alpha}{2\mu_0}B^2L. \quad \ldots\ldots \quad (9)$$

for points P and P'.

The total increase in the distance PP' is hence

$$\Delta L = \Delta L_{\text{elast.}} + \Delta L_{\text{mag.}},$$

from which formula (2) is derived.

Practical application of equation 2 to hum abatement

To be able to calculate the desired Young's modulus E_2 of the filling material for the gap, in addition to the given values of L, δ, O_1 and E_1, we must also know the values of a and S_1. These latter values must be found by measurement. Firstly, the amplitude of the vibration in the direction PP' of a coil core of the type shown in fig. 4 but without an air-gap is measured for a known magnetic induction, equal to the value which occurs in practice. In this case from (2) we have:

$$(\Delta L)_{\delta=0} = \frac{B^2}{2\mu_0} La \quad . \quad . \quad . \quad . \quad . \quad (10)$$

from which the value of a for the silicon iron used can be derived. The amplitude is then measured for a coil as in fig. 4 with air-gap but without filling material. Here, $E_2 = 0$, and consequently (2) becomes:

$$(\Delta L)_{E_2=0} = \frac{B^2}{2\mu_0}\left[aL - \frac{O_1}{S_1} + \frac{L}{E_1}\right] . \quad . \quad . \quad (11)$$

The value of S_1 calculated from (11) and of a from (10) are substituted in (2); then, putting $\Delta L = 0$, we have:

$$S_2 = E_2\frac{O_2}{\delta} = \frac{O_1}{La}\left(1 - \frac{E_1 O_1/L}{S_1}\right) - S_1, \quad . \quad (12)$$

and E_2 can now be calculated. A material of roughly this value can be chosen as filler, an additional correction being available in the choice of the cross-section $O_2 (\neq O_1)$ of the filling.

The arrangement that was used for making the measurements is shown schematically in *fig. 5*. The ballast under investigation is suspended on rubber bands, so that it can move freely. It is located between the probes of two vibration pick-ups of the type PR 9261 (formerly GM 5526). The vibration pick-ups, which are of the electrodynamic type, are enclosed in boxes of high-permeability alloy, to screen them from the leakage field from the ballast. The arrangement with two pick-ups represents a simple way of eliminating the effects of vibrations of the building. Thanks to the symmetry of the ballast under investigation, if the vibration pick-ups are in contact with symmetrically disposed points on the iron core, they will give voltages of the same phase and amplitude, since such points vibrate in phase and with the same amplitude (in opposite directions). Disturbances from outside, on the other hand, have an equal but opposite effect on the two vibration pick-ups, and may therefore be effectively eliminated by averaging the voltages delivered by the two pick-ups (M in fig. 5).

The electrodynamic vibration pick-up functions in such a way that the mean voltage is a measure of the velocity amplitude at the points investigated. What we wish to know, however, is the amplitude of the displacement, and it is therefore necessary to integrate the voltage obtained. This is done with the aid of an apparatus of the type PR 9250 (formerly GM 5522). The voltage from the latter (I in fig. 5) is applied to a cathode-ray oscilloscope via a selective amplifier (which passes only a narrow frequency band at 100 c/s) and via an electronic switch. The current in the coil of the ballast investigated is likewise supplied to the oscilloscope via the second channel of the electronic switch, this making it possible to see whether an expansion or a contraction occurs at the point investigated, at moments when the magnetic induction is at a maximum. To calibrate the apparatus, the amplitude on the oscilloscope screen is compared with a known alternating voltage of 100 c/s (not of the usual frequency of 50 c/s, because of the selective amplifier.)

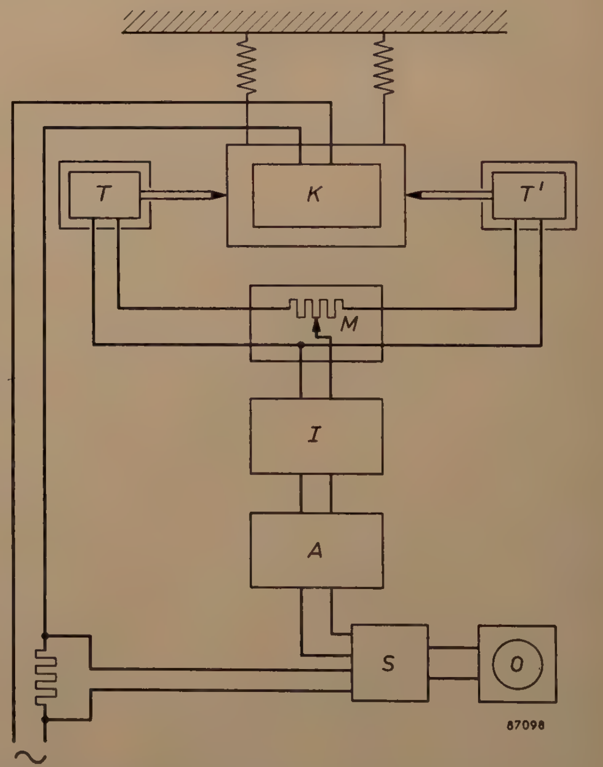


Fig. 5. Apparatus for measuring the vibration amplitude of an elastically suspended ballast K. T and T' electrodynamic vibration pick-ups, whose moving parts have a mass of about 10 grams. The pick-ups exert a constant pressure on the iron core of the ballast of about 8 N (1 newton is roughly equivalent to the weight of 100 grams). The greatest acceleration that can be measured, therefore, amounts to about 800 m/sec². The output voltages from the two pick-ups, added via the potentiometer M, are integrated (integrator I), amplified (in A) and displayed alternately with the current on a cathode-ray oscilloscope O, via an electronic switch S.

With the aid of the arrangement described, the vibration amplitudes (of the order of 0.1 micron or 10^{-7} m) occurring with symmetrical iron cores can be measured with an accuracy of 3×10^{-9} m. In the case of asymmetrical cores, where disturbances from the outside cannot be eliminated in the manner described, the accuracy of measurement is determined purely by the background disturbances. Only by working during quiet periods (at night for example) can an accuracy be obtained approximating to that which can be attained with symmetrical cores.

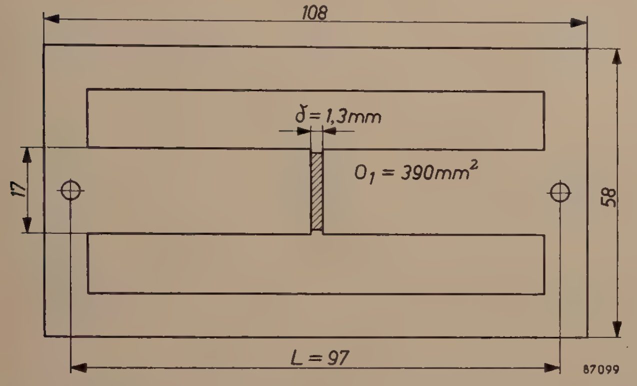


Fig. 6. Dimensioned diagram of the symmetrical core dealt with in the numerical example. The cross-section of the centre limb was $O_1 = 390$ mm².

To conclude this section we may take a numerical example relating to a choke with the dimensions shown in *fig. 6* and with $E_1 = 2 \times 10^{11}$ N/m², for use with 40 W lamps, 225 V, 50 c/s. Since the points of attachment P and P' are very close to the outside edge, the required amplitudes ΔL are virtually identical to the amplitudes measured at this outside edge.

Measurements on a core without air-gap gave $(\Delta L)_{\mathrm{mag}} = 0.108$ microns for $B = 1.25$ Wb/m², so that $a = 1.8 \times 10^{-12}$ m²/N. Measurements on a core with an air-gap, but without filling in the gap gave $(\Delta L)_{E_2=0} = -0.096$ microns for $B = 1.25$ Wb/m², from which it follows that $S_1 = 4.9 \times 10^{10}$ N/m². Substitution of the values found for a and S_1 in (12) gives $E_2 O_2/O_1 = 1.4 \times 10^9$ N/m². Such a low value for E_2 cannot, of course, be obtained with metals ($E \approx 10^{11}$), but can be obtained with materials like pressed insulating paper.

Asymmetric core with air-gap

In mass-production, choke coils of symmetrical design are sometimes less desirable for reasons of economy. A design frequently encountered in practice, is that shown in *fig. 7a*. Here we must take into account movements of P and P' (again the points of attachment) vertical to PP' as well as parallel to it.

Movements along PP' can be treated in exactly the same way as with the symmetrical construction. Thus, here also the ideal value can be found for Young's modulus of the gap-filling material. Clearly, S_2 will be appreciably smaller in this case than with the construction shown in fig. 4. The ideal filling does not, however, eliminate movements of P and P' at right angles to the line PP'; for the distance PP' remains the same, but the limb AA' expands owing to the preponderant magnetostriction. This means that bending will occur as indicated by the dotted line (which represents the centre line) in fig. 7*a* [8]).

It can be seen that the distance from PP' to the centre of gravity of the whole (which lies somewhere on MM') can vary. Moreover, the centre of gravity is not rigidly fixed, since the coil is not able to execute exactly all the movements made by the parts it encloses. This effect certainly cannot be ignored: the mass of the coil forms roughly half the total mass and must thus be taken into account; also the attachment of the coil is not completely free from play. On the average, the play is greater than the vibration amplitude of the core, so that it is a matter of chance where coil and core touch. The

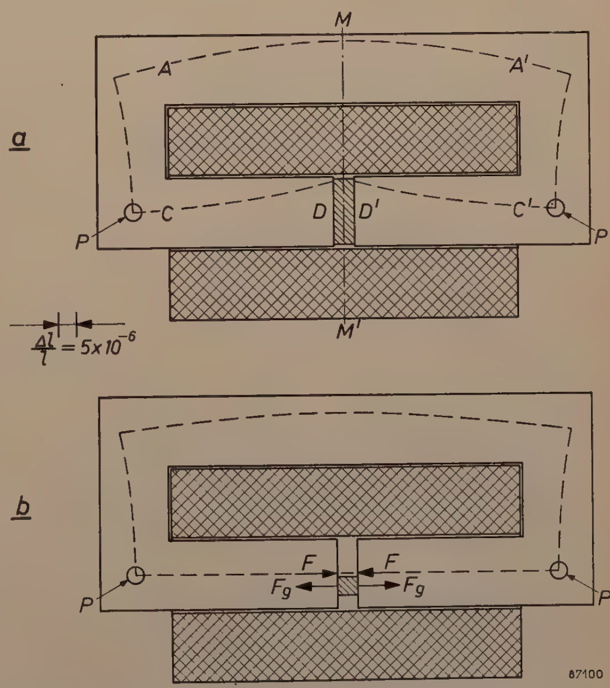


Fig. 7. Deformation pattern of a choke with an asymmetrical core. *a*) With the air-gap DD' uniformly filled. The distance PP' remains constant, but owing to the bending of the limbs AA' and CC' the centre of gravity is displaced along the line MM'. *b*) With the air-gap DD' asymmetrically filled. The bending of limb CC' is eliminated.

[8]) The shape of the dotted line was derived from measurements on a ballast, whose coil consisted of two parts, one on each half of the limb bearing the coil. The iron core was thus accessible at the air-gap.

mutual movement of coil and core differs, therefore, from case to case, and has the unfortunate consequence, that the amplitude of the vibration of P and P' perpendicular to the line PP' has no constant, calculable value. It is only possible to give a maximum value.

The shape of the dotted line in fig. 7*a* suggest that one means of countering the mutual movement of core and coil might be by applying a bending torque, proportional to the force of magnetic attraction between the poles D, D', acting so as to exactly eliminate the bending of the halves of the limbs, described above (there is no objection to limb AA' bending). This has been found to be possible by filling the air-gap, not uniformly with the material with the exact Young's modulus, required by the condition that $\Delta L = 0$, but by filling it with a material with a greater Young's modulus on the outside and a material with a smaller Young's modulus on the inside. The filling can be adjusted so that the condition $\Delta L = 0$ is still satisfied, but since the point of application of the resulting force F_g is displaced towards the outside, this force, together with the force of magnetic attraction F of the poles (acting in the centre line) forms the desired bending torque. The optimum arrangement is best found by experiment. With a choke in which $\Delta L = 0$ and in which the artifice just described has been applied, the coil will not be able to move at right angles to PP', irrespective of where the points of contact are between coil and core (fig. 7*b*).

Naturally, if P and P' are chosen as points of attachment, the movement of the centre of gravity perpendicular to PP' will not be completely eliminated, for AA' will still be bent. It might seem worth while to see whether any improvement can be gained from a slight over-compensation of the movement of the coil with respect to the core; experience shows, however, that even with the best attachment of the coil to the core obtainable in mass production, such a process again leads to excessive differences in the resulting vertical component of the individual coils. Fortunately, however, the reduction in hum attained with the compensation described is already very satisfactory.

Cores without air-gaps (transformers)

Where the mains voltage is low (110 or 127 V) the ballasts usually consist of leakage transformers, i.e. of a transformer in combination with a choke. Sometimes, however, small transformers are also used separately. This is the case, for example, with type "TL"S, where normally an incandescent lamp functions as ballast [9]), but where, in addition, a transformer is necessary if a mains voltage of 220 V is not available.

Since these transformers do not possess air-gaps, the process outlined above for combating hum cannot be applied.

In principle two solutions are feasible:

[9]) W. Elenbaas and T. Holmes, Philips tech. Rev. **12**, 129-135, 1950/51.

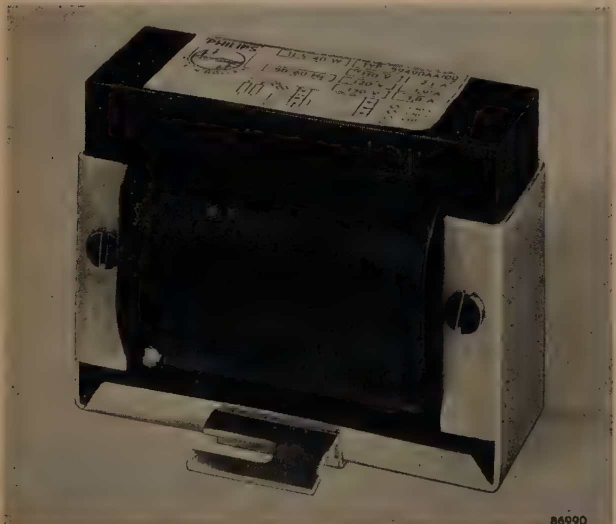


Fig. 8

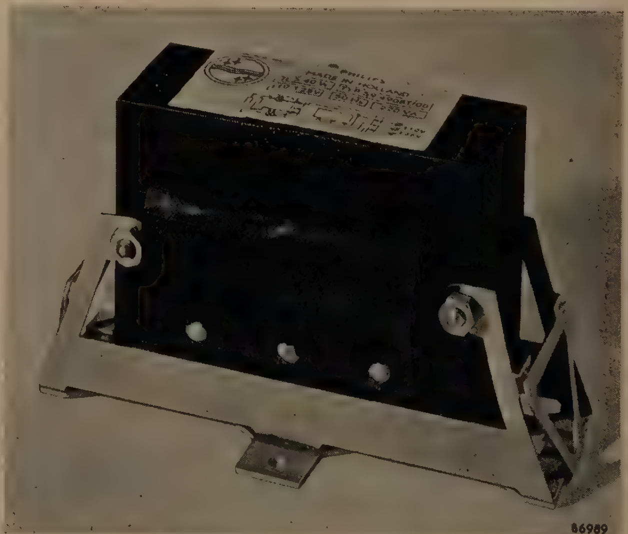


Fig. 9

Fig. 8. Transformer in a rectangular bracket with flexible corners. The bracket is screwed to the mounting surface through the middle of the horizontal connecting piece.
Fig. 9. Transformer in a rigid double bracket with convergent arms.

a) The use of transformer iron which shows hardly any magnetostriction.
b) Attachment of the transformer exactly at its centre of gravity.

In practice neither solution is feasible. "Free" suspension (on springs) is also out of the question.

It has however been found that if only two points suitable for attachment can be found on the iron core, which vibrate along the same line in phase and with equal and opposite amplitudes, a bracket can be constructed, which while being sufficiently firm, does not transmit vibrations to the ceiling or wall.

Two types of bracket have been developed, both for symmetrical transformers; these are shown in *figs. 8* and *9*.

The type shown in fig. 8 is very simple. The body and arms of the right-angled bracket are reinforced with flanges; the corners are purposely left flexible. The core of the transformer may be assumed here to be the same as that of the choke in fig. 4, but without an air-gap. Here, apart from the symmetry of the vibrations mentioned earlier, use is also made of the shape of the deformed iron core.

The shape of the core, as measured in the case of the transformer for which this bracket was designed (for $B_{max} = 1.25$ Wb/m²), is given in *fig. 10*. The central parts of the yokes of the core, between X and Y or X' and Y', bend but little and move almost

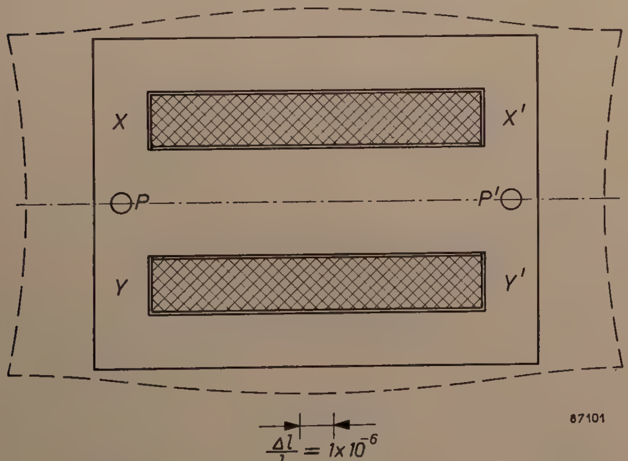


Fig. 10. Deformation pattern of a transformer with symmetrical core, of the type shown in figs. 8 and 9. The dotted line indicates the outline of the deformed core.

parallel to each other. If the arms of the bracket are clamped sufficiently firmly to the core at these places (and this is possible even with a relatively low coefficient of friction), they will then be displaced parallel to each other. This means that the connecting piece of the bracket will be stretched without bending. The centre of the connecting piece thus remains stationary and can function as the point of attachment of the whole transformer to a wall, etc.

The second type, a convergent double-bracket, is represented diagrammatically in *fig. 11*, which shows only the elastically neutral line of the body

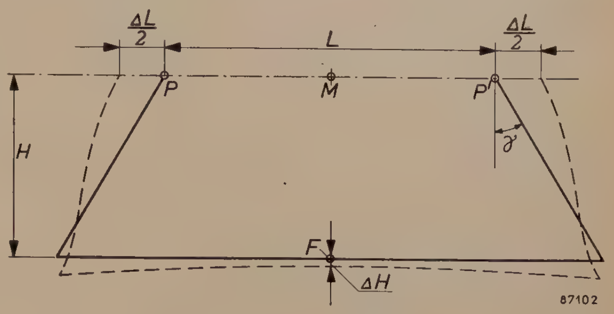


Fig. 11. Diagram of the deformation of the halves of the bracket shown in fig. 9.

and the arms (for one half of the bracket). The way in which the ends of the arms of the two halves of the bracket are attached to the transformer core can be seen in the photograph, fig. 9. The centre F of the connecting piece of each half of the bracket is screwed on to the wall to which the transformer is to be affixed. In a bracket of the type shown, which has rigid corners, substantial bending of the connecting piece will occur, and in general there will be a displacement ΔH of F with respect to the centre of gravity M of the transformer (which, on account of the relatively low mass of the bracket, practically coincides with the centre of gravity of the whole system). However, if the bracket has suitable dimensions $\Delta H = 0$, so that no vibration will be transmitted to the wall, i.e. no hum will be audible.

At moments when the induction B is a maximum, points P and P' of the core in fig. 11 are both displaced towards the outside through a distance $\frac{1}{2}\Delta L$. Provided the transformer is sufficiently firmly attached to the bracket, the extremities of the arms move with P and P'. The shape then assumed by the neutral line, is indicated (exaggerated) by the dotted line in fig. 11. Calculation of the elastic deformation gives the following expression for the displacement ΔH from the centre F:

$$\frac{\Delta H}{\Delta L} = \frac{\frac{3}{4}\left(\frac{L}{H}\right)^2 \cos\gamma + \left(\frac{L}{H}\right)\sin\gamma - \sin\gamma\,\mathrm{tg}\,\gamma - \frac{I_1}{I_2}\mathrm{tg}\,\gamma}{4\left(\frac{L}{H}\right)\cos\gamma + 4\sin\gamma + \frac{I_1}{I_2}}. \quad (13)$$

I_1 and I_2 are the linear moments of inertia or second moments of area [10]) of the cross-sections of the body and the two arms respectively; for the significance of L, H, γ see fig. 11.

[10]) For the relevance of the second moment of area $I = \int y^2 \mathrm{d}A$ of a cross-section to the elastic strain, the reader is referred to text-books on elasticity. For a rod of rectangular cross-section of breadth b and length h, the axis running centrally across the area, parallel to b, $I = bh^3/12$. I is expressed in inch⁴ or m⁴ (here, the latter).

It follows from formula (13) that ΔH can be either positive or negative and can also be made equal to zero. For the latter, the condition:

$$\frac{I_1}{I_2} = \tfrac{3}{4}\left(\frac{L}{H}\right)^2 \frac{\cos^2\gamma}{\sin\gamma} + \left(\frac{L}{H}\right)\cos\gamma - \sin\gamma. \quad . \quad . \quad (14)$$

must be satisfied. For given values of L, H and γ the desired value of I_1/I_2 can be read off from *fig. 12*, in which the ratio I_1/I_2 is plotted against L/H (both scales logarithmic), for various values of γ.

It should be remarked that equation (14) is valid only when the natural frequencies of the equipment are high compared with the vibration frequency of the transformer (100 c/s). This condition can be easily satisfied, since a condition is imposed only on the ratio I_1/I_2, while I_1 or I_2 may still be chosen at will.

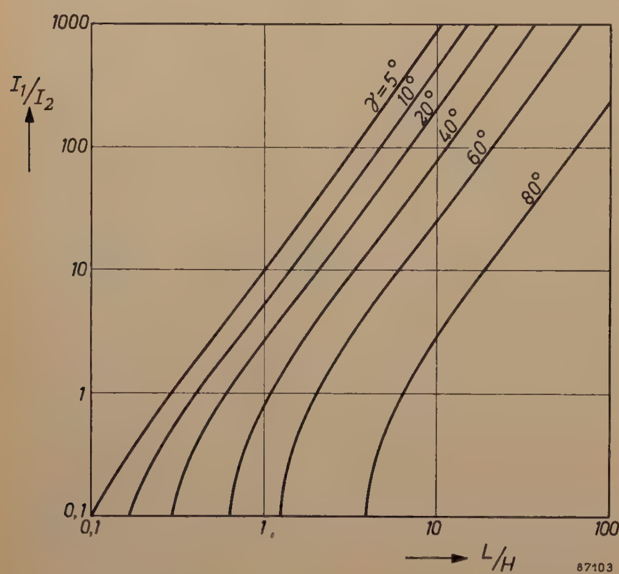


Fig. 12. Graph giving the desired value of I_1/I_2 as a function of L/H, for various values of the angle γ (see fig. 11).

Measures to reduce rustle

We have seen that hum is the consequence of vibrations set up in the core of the ballast, which are transmitted to the wall.

The core can also vibrate freely (to be compared with a freely vibrating tuning fork). Owing to the large cross-section, these vibrations are of high frequency (several kc/s). Moreover, since the core is built up of laminations, they are strongly damped. The vibrations are produced in an irregular manner by the small current peaks described earlier, which are superimposed on the approximately sinusoidal current. Owing to the indeterminacy of the phases and the spread in the frequency of these vibrations, they give rise to a noise spectrum known as rustle.

It is evident that rustle will occur only with choke coils. With transformers the magnetization is substantially sinusoidal owing to the low impedance of the mains shunting the primary winding.

At the frequencies involved with rustle, the transmission of vibrations to the walls is found to be very slight [11]). This is fortunate, since the mounting methods for countering hum discussed above, can generally offer no solution: the calculations made in connection with these methods were based on the assumption that the forces and the deformations they produce are in phase. This holds well enough at low frequencies, but not at the high frequencies associated with rustle, which lie within the range of the natural frequencies of the equipment.

The chief problem, therefore, is to prevent direct emission of sound. This can be best achieved by surrounding the ballast by a closed box; however, other problems then arise. In connection with the measures against hum, the equipment should be mounted in the box only at vibration-free points (nodal points), yet to obtain sufficient dissipation of the heat developed the box should be packed with some heat-conducting mass. The filling necessarily forms a mechanical contact between the walls of the box and the vibrating parts of the equipment, and must therefore be not only a good conductor of heat but also a good insulator against vibrations. For vibrations in the frequency range of rustle it would be sufficient that the equipment nowhere makes direct contact with the box. For hum frequencies, which are more difficult to insulate, the fact that on opposite sides of the nodal points the displacement is in opposite directions has the result that the net transmission of vibrations by the filling to the box and thence to the mounting surface is sufficiently attenuated.

A common solution is to float the ballast, as it were, in a bituminous mass. To improve heat conduction, sand is frequently added to the bitumen. A difficulty here is that whilst the bitumen filler should preferably be soft and pourable at a temperature high above the working temperature of the coils, the electrical insulation cannot withstand such a temperature during manufacture. For this reason, a filling material has to be chosen which softens at relatively low temperatures, with the attendant risk that sooner or later the ballast will sink and come into contact with the metal wall of the box, greatly to the detriment of the acoustical insulation.

A good solution to the problem has been found by filling the space in the box with corrugated aluminium foil (silver paper). If a suitable "filling factor" be chosen, the heat conductivity is quite as good as or better than that of the bitumen-sand mixture.

11) W. Elling Bestimmung mechanischer Eingangsimpedanzen, Acustica 4, 396-402, 1954.

Vibrations in the frequency range of rustle are adequately damped, since they must pass many contact points between the sheets of foil, before reaching the wall of the box.

In conclusion, we shall briefly note the behaviour of the filling with respect to the hum vibrations of the iron core. Initially, of course, points on all parts of the iron core will come into contact with the aluminium. Some of these points vibrate with relatively large amplitudes, others with small amplitudes. Where the hum amplitudes are large, the elastic limit of the sheets of foil will be exceeded, so that these sheets will be plastically deformed by the vibrating iron core. In this way room will be made for vibration, the final result being, that the aluminium filling will only be in contact with the iron core where the amplitude of the hum is small or non-existent and the remainder of the core will be left free. This is just what we want, for now the low frequency vibrations are not transmitted via the box to the wall or ceiling.

Proof of this explanation is furnished by a simple experiment. If the whole apparatus (i.e. complete with box) is put in another position, then owing to the change in direction of the gravitational force, the sheets of foil will come into contact with quite different points on the iron core. A certain increase in hum can indeed often be heard, but the latter rapidly falls once more to its former low and acceptable level.

Summary. The origin of troublesome noises produced by fluorescent lighting installations lies in the ballasts. The latter vibrate with twice the mains frequency (producing hum) and also execute free vibrations of high frequency (rustle). Both forms of vibration result from the alternating elastic and magnetostrictive deformations due to the variations of current. High-frequency vibrations are directly transmitted to the air, low-frequency vibrations via the wall or ceiling. In ballasts with air-gaps (chokes), hum can be countered by filling the gap with a suitable material (mutual compensation of magnetostrictive and elastic deformation at the points of attachment). In ballasts without air-gaps (transformers) transmission of vibrations to the wall can be avoided by use of suitably designed brackets. Rustle can be combated by enclosing the ballast, surrounded by a sound-insulating material, in a box. The insulating material must however beheat conducting; aluminium foil has been found to answer the purpose.

THE GYRATOR, AN ELECTRIC NETWORK ELEMENT

by B. D. H. TELLEGEN.

621.372.2

There is a traditional antithesis between the practical mind and the mind inclined to pure scientific pursuits: the former is mainly interested in knowledge that can be put immediate to use, while the latter ponders fundamentals and tries to introduce greater clarity and generality by rigorous considerations which the practical man probably finds superfluous. The invention of the gyrator by Prof. Tellegen is a striking illustration of the service that can be rendered to engineering by the pure scientific approach: originally regarded as a hypothetical possibility, a possibility that had to be recognized for the sake of completeness, the gyrator has subsequently and perhaps rather surprisingly become a reality in the world of microwaves.

Wide use is made in electrical engineering of networks composed of resistors, coils and capacitors, the latter being known as network elements. In use, there is conversion and exchange of energy within and between these network elements. In resistors electrical energy can be converted into heat; in coils and capacitors energy can be stored and later released. The networks are provided with terminal pairs allowing an exchange of energy to take place with the exterior. A terminal pair consists of two conductors between which a voltage of instantaneous value v can exist and through which a current of instantaneous value i can flow (see *fig. 1*). The energy

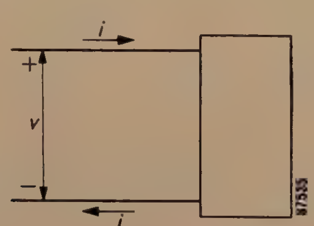


Fig. 1. Network with one terminal pair.

supplied to the network via the terminal pair in a time dt is $iv\mathrm{d}t$; it can be positive or negative. The network determines relations between the voltages and currents at the terminals, the number of relations being equal to the number of terminal pairs. If for example the network consists of one resistor R, it has one terminal pair and the relation is $v = Ri$.

Networks made up of resistors, coils and capacitors

For the user networks are characterized primarily by the relations that hold between the terminal voltages and currents. As long as the relations between these voltages and currents are the ones desired, the actual composition of the network is, for the user, of secondary concern. It is therefore of great importance to know what sets of such relations are possible with networks consisting of resistors, coils and capacitors. All the possible sets of relations form an arsenal from which the user can take his choice.

In order to build up the whole arsenal, i.e. a complete list of the possible sets of relations, let us first look for general properties of the relations; we can then try to demonstrate that, for each set of relations having these properties, it is possible to design a network of resistors, coils and capacitors for which the said relations are valid. This can be put in another way by saying that we shall try to find necessary and sufficient conditions for the sets of relations if they are to be realizable by networks composed of resistors, coils and capacitors.

General properties of the relations are that they consist of *linear* differential equations relating the terminal voltages and currents, and that the coefficients of the equations are *constant*, i.e. not dependent on time but determined only by the magnitude of the elements composing the networks. Furthermore, the networks contain no source of energy; they are said to be *passive*. Certain properties of the coefficients of the differential equations can be deduced from the fact that the networks are passive. Let us consider, for example, a network with one terminal pair, the properties of which are governed by a differential equation of the first order, so that we can write:

$$a\frac{\mathrm{d}i}{\mathrm{d}t} + bi = c\frac{\mathrm{d}v}{\mathrm{d}t} + dv, \quad . \; . \; . \quad (1)$$

where a, b, c and d are constants. From the passivity of the network it can be shown that a, b, c and d all have the same sign.

This can be demonstrated as follows. If we short-circuit the network, in other words if we keep $v = 0$, the current will be determined by:

$$a\frac{\mathrm{d}i}{\mathrm{d}t} + bi = 0.$$

The solution of this differential equation is:

$$i = C\, e^{-\frac{b}{a}t}$$

where C is the constant of integration. In consequence of the passivity, the current in the short-circuited network cannot rise indefinitely and hence a and b have the same sign. It follows in a similar way, by considering the network to be open-circuited, i.e. by keeping $i = 0$, that c and d have the same sign.

For direct current and voltage equation (1) simplifies to

$$bi = dv,$$

so that b/d represents the DC resistance of the network. This must be positive on account of the network's passivity, and hence b and d also have the same sign.

For convenience we shall group the three properties together and say that the sets of relations characterizing networks made up of resistors, coils and capacitors are linear, constant, passive. As far as networks with one terminal pair are concerned, these properties are not only necessary but also sufficient, for Brune has demonstrated [1]) that *any* linear constant passive relation can be realized by a network made up of resistors, coils and capacitors. In particular, it can be demonstrated that any relation having the form of equation (1) can be realized by a network consisting either of two resistors and a coil (if $a/c > b/d$), or of two resistors and a capacitor (if $a/c < b/d$). The two networks are shown in *fig. 2*.

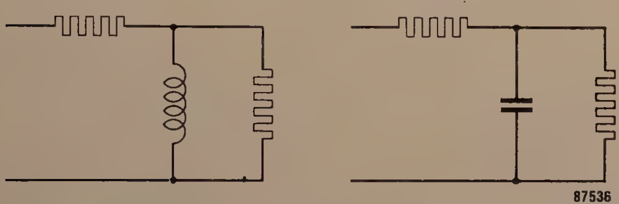


Fig. 2. Networks with one terminal pair and of the first order.

For any network consisting of resistors, coils and capacitors and having more than one terminal pair a set of relations holds that, besides being linear constant passive, has the property known as reciprocity. To explain this property we shall take a network with two terminal pairs and express the two terminal voltages in terms of the two terminal currents; for the present purpose it will be convenient to write the relationships in complex form, rather than use the instantaneous values. Accordingly, we obtain:

$$\left.\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2, \\ V_2 &= Z_{21}I_1 + Z_{22}I_2, \end{aligned}\right\} \quad . \; . \; . \; . \; . \quad (2)$$

where I_1, I_2, V_1 and V_2 represent the complex values of the terminal currents and voltages. If we assume the sign convention for currents and voltages as indicated in *fig. 3*, it can be shown that always

$$Z_{21} = Z_{12}. \quad . \; . \; . \; . \; . \; . \; . \; . \quad (3)$$

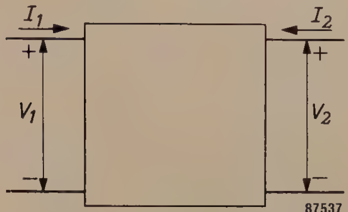


Fig. 3. Network with two terminal pairs.

Expressing the current I_1 through the first terminal pair and the voltage V_2 across the second terminal pair in terms of the two other terminal quantities, we find from equations (2):

$$\left.\begin{aligned} I_1 &= \frac{1}{Z_{11}}V_1 - \frac{Z_{12}}{Z_{11}}I_2, \\ V_2 &= \frac{Z_{21}}{Z_{11}}V_1 + \frac{Z_{11}Z_{22} - Z_{12}Z_{21}}{Z_{11}}I_2. \end{aligned}\right\} \quad . \; . \quad (4)$$

The coefficient of V_1 in the second equation above is now equal and opposite in sign to the coefficient of I_2 in the first equation. The relation (3) between two coefficients of equations (2) and the relation just discussed between two coefficients of equations (4) are known as the reciprocity relations.

For networks with more than one terminal pair Bayard and others have demonstrated [2]) that any linear constant passive set of relations having the reciprocity property can be realized by a network consisting of resistors, coils and capacitors. Hence the said properties of sets of relations are sufficient as well as necessary.

Our arsenal of all possible sets of relations capable of realization by networks made up of resistors, coils and capacitors is thus complete.

Linear constant passive systems

The result obtained above is not entirely satisfactory, however. We have based our considerations on the resistors, coils and capacitors that had their origin in the laboratory. Electrical engineering has simply accepted these elements for constructing networks. There is something arbitrary and fortuitous about all this. Must we necessarily use these

[1]) O. Brune, J. Math. Phys. **10**, 191, 1931.

[2]) M. Bayard, Bull. Soc. franç. Elect. **9**, 497, 1949. Also B. D. H. Tellegen, J. Math. Phys. **32**, 1, 1953, which gives further references to the literature.

elements for the building-up of networks? Should an affirmative answer be found to this question, we would be faced with another one: are other elements conceivable, elements that have not been found in the laboratory?

In order to investigate these questions we must not proceed from the basis of the conventional network elements and the networks composed of them but approach the problem from another direction. We must consider "black boxes", systems with terminal pairs that are characterized only by the sets of relations existing between the terminal voltages and currents without concerning ourselves with what is inside them. In what follows we shall confine ourselves to linear constant passive systems, by which we mean systems characterized by linear constant passive sets of relations. We could of course subject ourselves to less stringent limitations: we could also consider systems containing energy sources (active systems, such as amplifiers), systems characterized by linear equations with variable coefficients (variable systems), or systems characterized by non-linear equations (non-linear systems). However, such systems have properties that depart appreciably from those of the networks considered above, and are of a more complex nature. We shall therefore leave them aside.

Let us now try to find the "simplest" kinds of linear constant passive system. These simplest systems we shall call network elements, *by definition*. In this way we shall try to arive at a set of network elements such that any linear constant passive system can be realized as a network composed of them. We can regard such a set of network elements as a complete set.

In order to give a meaning to the epithet "simplest", we shall have to classify our "black boxes". We shall do so according to the number of terminal pairs, the order of the differential equations characterizing them, and whether or not they are able to dissipate electrical energy, i.e. to transform it into heat. This leads, as a first step, to the investigation of systems with one terminal pair, of zero order and involving no dissipation. For such systems the power supplied must be zero at any instant, i.e. $iv = 0$; therefore either $i = 0$ (open terminal pair) or $v = 0$ (short-circuited terminal pair). This does not produce a network element. In accordance with the threefold classification, our next step will be to investigate three classes of systems, namely (1) systems with one terminal pair, of zero order and involving dissipation, (2) systems with one terminal pair, of the first order and involving no dissipation, and (3) systems with two terminal pairs, of zero order and involving no dissipation.

(1) Systems with one terminal pair, of zero order and *involving dissipation*

These systems are characterized by an equation having the form:

$$v = Ri, \quad R > 0 \; . \; . \; . \; . \; . \; . \quad (5)$$

The fact that R is positive is a consequence of the passivity of the network, this requiring that iv should be positive.

(2) Systems with one terminal pair, of the *first order* and involving no dissipation

Systems with one terminal pair and of the first order are characterized by an equation of the form of (1). It can be shown that this represents a system involving no dissipation either if $b = 0$ and $c = 0$ or if $a = 0$ and $d = 0$. Hence these systems are of two kinds. The first kind is characterized by an equation of the form:

$$v = L\frac{\mathrm{d}i}{\mathrm{d}t}, \quad L > 0, \; . \; . \; . \; . \; . \quad (6)$$

and the second kind by an equation of the form:

$$i = C\frac{\mathrm{d}v}{\mathrm{d}t}, \quad C > 0 \; . \; . \; . \; . \; . \quad (7)$$

The above may be derived as follows. The system to which (1) is applicable has an impedance of

$$Z = \frac{j\omega a + b}{j\omega c + d}.$$

In order that there should be no dissipation, the real part of Z must be zero at all frequencies. It follows that $ac = 0$ and $db = 0$. Since $a = 0$ and $b = 0$ means that $Z = 0$, and $c = 0$ and $d = 0$ means that $Z = \infty$, these solutions may be disregarded. Thus either c and b, or a and d, must be zero, and these two cases produce equations (6) and (7), respectively. Since the passivity of the network requires that a, b, c and d should all have the same sign (see above), it further follows that $L = a/d$ and $C = c/b$ are both positive.

Before examining the third class of simplest systems we shall take a closer look at the three systems, (5), (6), and (7), already found. As stated above, we shall regard them as network elements. The latter, then, are defined by equations, not by the physical means required to realize them. For example, we cannot infer from the equation $v = L\,\mathrm{d}i/\mathrm{d}t$ that it describes a "coil". A superconductor has the same equation when we take the effect of the mass of the conducting electrons into account. The kinetic energy of these electrons then takes the place of the magnetic energy of the coil.

The three network elements, so defined, are "ideal"; they can only be realized approximately. For example, (5) is approached by a resistor with low stray capacitance and low self-inductance (6) by a low-loss coil with a low stray capacitance, and (7) by a low-loss capacitor with a low self-inductance (hence the symbols R, L and C used in the equations).

The way in which the network elements are realized is only of secondary importance to the user; it is the external properties given by (5), (6) and (7), that are of primary interest. This remark is similar to that made earlier regarding networks: the user has little interest in their internal composition.

(3) Systems with *two terminal pairs*, of zero order and involving no dissipation

For these systems the total power supplied via the terminal pairs is zero at any instant, so that

$$i_1v_1 + i_2v_2 = 0,$$

where i_1, i_2, v_1 and v_2 represent instantaneous values of the terminal currents and voltages, the sign convention for these quantities being assumed to be in accordance with fig. 3. This condition results in two kinds of systems, namely those characterized by equations of the form:

$$\left.\begin{aligned} i_1 &= -ni_2, \\ v_2 &= nv_1, \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (8)$$

and those characterized by equations of the form:

$$\left.\begin{aligned} v_1 &= -si_2, \\ v_2 &= si_1, \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (9)$$

Equations (8) and (9) may be derived as follows. If, for a two-terminal-pair system of zero order, we express the two terminal voltages in terms of the two terminal currents, we can write the equations as:

$$\left.\begin{aligned} v_1 &= a_{11}i_1 + a_{12}i_2, \\ v_2 &= a_{21}i_1 + a_{22}i_2. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (10)$$

It follows from this that:

$$i_1v_1 + i_2v_2 = a_{11}i_1^2 + (a_{12} + a_{21})\, i_1i_2 + a_{22}i_2^2. \quad \ldots \quad (11)$$

In order that (10) should represent a system involving no dissipation, (11) must be zero for all values of i_1 and i_2; hence $a_{11} = 0$, $a_{12} + a_{21} = 0$, and $a_{22} = 0$, which results in (9).

Starting from the equations that express i_1 and i_2 in terms of v_1 and v_2 we also arrive at (9). However, it is also possible to conceive systems in which v_1 and v_2 cannot be expressed in terms of i_1 and i_2, or vice versa. This is the case for systems that are characterized by a relation between i_1 and i_2 and a relation between v_1 and v_2, making it impossible to chosse i_1 and i_2 or v_1 and v_2 as independent variables. In order to investigate these systems we can start from equations expressing i_1 and v_2 in terms of v_1 and i_2, or vice versa. We then arrive at (8).

We shall regard the systems thus found, (8) and (9), as two further network elements. System (8) is called an ideal *transformer* and we have proposed [3]) the name ideal *gyrator* for system (9). Observations similar to those made above on (5), (6) and (7) may also be made on these two systems.

An approximation to the ideal transformer is given by two tightly coupled low-loss coils of large self-inductances. This is in fact the way in which the concept first arose. The theoretical route by which we have now arrived at the same concept shows why we should regard the ideal transformer as a separate network element, defined, in fact, by (8). From the remarks on reciprocity in connection with equations (4) we see that the ideal transformer also possesses this property. The ideal gyrator does *not* possess the property of reciprocity, as will be clear from a comparison of (9) with (2) and (3). It is therefore impossible to obtain an approximation to it by combining resistors, coils and capacitors, because such a combination always has the property of reciprocity, as we saw above. For the realization of the gyrator other physical means are necessary. One way of realizing it, making use of gyromagnetic effects in ferromagnetic materials, will be described in a further article in this Review [4]).

The addition of the gyrator to the set of network elements has extended the list of possible sets of relations between terminal voltages and currents. This extension was made possible by the fact that we started from linear constant passive systems as such, without imposing on them the condition of reciprocity. The fact that system (9), while being linear constant passive lacks the property of reciprocity, shows that the latter is not a consequence of linearity, constancy and passivity, and that imposing it does indeed constitute a limitation of the possibilities.

Some properties of the gyrator

The ideal gyrator has the property of "gyrating" a current into a voltage, and vice versa. The coefficient s, which has the dimension of a resistance, we call the gyration resistance; $1/s$ we call the gyration conductance. We shall represent the gyrator in circuit diagrams by the symbol shown in *fig. 4*.

The following properties of the ideal gyrator can be easily derived from (9).

---

[3]) B. D. H. Tellegen, Philips Res. Rep. 3, 81, 1948.
[4]) H. G. Beljers, Application of ferroxcube to uni-directional waveguides, to appear shortly in this Review.

If we have the output terminals open-circuited, i.e. $i_2 = 0$, the input terminals are short-circuited,

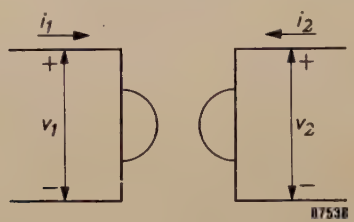


Fig. 4. Symbol for the gyrator.

i.e. $v_1 = 0$, and vice versa. If we connect an inductance L or a capacitance C across the output terminals, we find a capacitance L/s^2 or, in the second case, an inductance s^2C between the input terminals. In general, if we connect an impedance Z across the output terminals, we find an impedance s^2/Z between the input terminals. An impedance Z in series or in parallel with the output terminals has the same effect as an impedance s^2/Z in parallel or in series, respectively, with the input terminals (*fig. 5*).

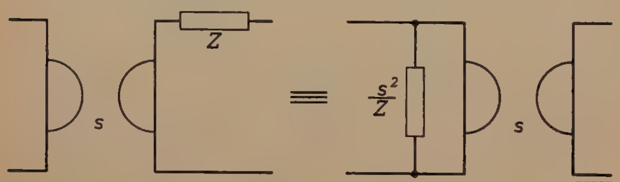


Fig. 5. An impedance in series with one terminal pair of an ideal gyrator is equivalent to another impedance in parallel with the other terminal pair.

Two ideal gyrators in cascade form an ideal transformer; an ideal gyrator and an ideal transformer in cascade form another ideal gyrator.

The combination of a gyrator with a resistor shown in *fig. 6a* gives a system to which the following apply:

$$\left.\begin{aligned} v_1 &= Ri_1 + (R-s)i_2, \\ v_2 &= (R+s)i_1 + Ri_2. \end{aligned}\right\} \quad \ldots \quad (12)$$

Thus an input i_1 gives rise to a voltage component $(R+s)i_1$ across the output; an output current i_2 gives rise to a voltage component $(R-s)i_2$ across the input. If $R = s$, the latter component is zero, so that the input voltage is independent of the output current and dependent only on the input current. In such a case we may say that the system transmits only in the direction from input to output, and not in the reverse direction. The system shown in fig. 6*b* has properties of a similar kind; this can be easily demonstrated by writing down the equations expressing the terminal currents in terms of the terminal voltages.

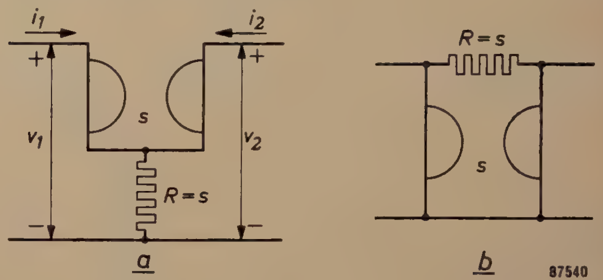


Fig. 6. Uni-directional networks, i.e. networks that transmit only in one direction.

No further elements remain to be added to the five ideal linear constant passive network elements: Oono and Yasuura have demonstrated [5]) that any linear constant passive system can be realized as a network composed of these five network elements. The five ideal network elements therefore form a complete set.

[5]) Y. Oono and K. Yasuura, Mem. Fac. Engng., Kyushu Univ. **14**, 124, 1954; also in Ann. Télécomm. **9**, pp. 73 and 109, 1954.

Summary. The writer considers the electric networks having one or more terminal pairs that can be built up from conventional network elements, namely resistors, coils and capacitors. The set of relations between terminal voltages and currents determined by such a network has properties that correspond to the linear, constant and passive nature of such networks and, furthermore, has the property of reciprocity. It has been demonstrated that, conversely, these properties are sufficient for any set of relation possessing them to be realizable by a network. The writer reverses this line of reasoning and raises the question: what (ideal) network elements must be introduced in order to make it possible to realize all linear constant passive systems. There are no grounds for including reciprocity amongst the properties imposed. It is then shown that, apart from the conventional elements — resistor, coil, capacitor, and the ideal transformer — a new network element has to be introduced; this new element has been christened ideal gyrator. Networks containing the new element generally lack the property of reciprocity. A brief sketch is given of some of the properties of the gyrator.

LEVITATION BY STATIC MAGNETIC FIELDS

531.51:538.12

Levitation of objects is a problem that has fascinated philosophers throughout the ages. A solution of this problem may have practical importance, e.g. for eliminating friction in rotation or horizontal displacement, or for avoiding chemical reactions between the object and its surroundings. With an eye to the practical importance of levitation we feel justified here in disregarding those aspects of it associated with magic, spiritualism and psychic phenomena, and confining ourselves to a few remarks concerning its physical realization and, more particularly, to a description of some informative experiments in this field.

The problem is that of compensating the gravitational force exerted on a body in such a way that, *in the absence of any mechanical contact,* it is held in stable equilibrium with respect to the earth. Probably the most spectacular solution is that where the compensating forces are generated by the body itself as reactive forces, in accordance with the rocket principle. A feature of the rocket solution is that it is in principle possible to levitate a body at *any* height above the earth. For every pound of weight lifted, however, an enormous power is required, so much indeed that using conventional fuels levitation cannot be maintained for more than a few minutes. The use of nuclear power may be calculated to raise the theoretically attainable maximum duration to about 250 days [1]).

When the gravitational forces exerted on the body are compensated by external forces, such as those exerted by electric or magnetic fields, levitation is possible only at relatively short distances from the equipment that generates these forces and which are themselves earthbound. In these cases, though, considerably less power is required and levitation can be maintained for an unlimited period. In this connection we would mention a spectacular experiment carried out in 1939, in which an aluminium disc of about a foot in diameter was made to float freely in an alternating magnetic field, generated by a special magnet fed with alternating current from the mains [2]). Here the gravitational force was compensated by the force exerted by the magnetic field on the electrons moving in the conductor. An analogous arrangement was employed some years ago for melting metals in vacuo without a crucible [3]) — a technique whose potentialities might prove to be important.

The essential feature of the above experiments is that the levitated body is kept in *stable* equilibrium. According to a well-known theorem, due to Earnshaw, this cannot be the case with a magnetic body floating freely (i.e. without any mechanical support) in a *constant* magnetic field [4]). It is evident, for example, that a small permanent magnet M_2 suspended below a large permanent magnet M_1 such that the gravitational force on M_2 is exactly compensated (*fig. 1*) cannot be in stable equilibrium. The slightest upward displacement of M_2 results in a greater magnetic attraction and M_2 continues to rise until it clings to M_1; any slight downward displacement of M_2 reduces the magnetic attraction and M_2 then continues to fall. This unstable equilibrium can, however, be turned into a stable one if the permanent magnet M_1 is replaced by an electromagnet, the excitation current of which is automatically regulated in accordance with the position of M_2. An elegant set-up for this purpose is to arrange that the position of M_2 controls the value of a capacitance, which in its turn controls the excitation of the magnet. In an experimental arrangement, a body weighing 210 kg (about 4 cwt) was levitated in this way [5]).

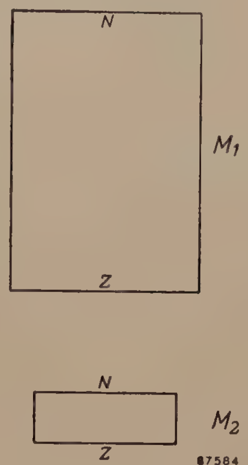


Fig. 1

The power required in the above case was remarkably low: only 1.3 watts per kg. However, levitation may be effected by a static magnetic field, consuming no energy at all. Earnshaw's theorem, according to which a magnetic body floating in a static magnetic field cannot be in a state of stable equilibrium, is invalid for substances with a relative

1) For this calculation, and more generally for a survey of the theoretical and practical possibilities of levitation, reference should be made to: A. H. Boerdijk, Technical aspects of levitation, Philips Res. Rep. **11**, 45-56, 1956.

2) B. D. Bedford, L. H. B. Peer and L. Tonks, Gen. El. Rev. **42**, 246-247, 1939.

3) E. C. Okress, D. M. Wroughton, G. Comenetz, P. H. Brace and J. C. R. Kelly, J. appl. Phys. **23**, 545-552, 1952.

4) S. Earnshaw, Trans. Cambr. Phil. Soc. 7, 97-112, 1842. See also: J. C. Maxwell, A treatise on electricity and magnetism, Clarendon, Oxford 1873, part I, pp 139-141;

5) H. Kemper, E.T.Z. **59**, 391-395. 1938.

magnetic permeability smaller than unity, as has been demonstrated by Braunbek [6]). Returning to the arrangement of fig. 1, it is possible to stabilize the equilibrium of the small permanent magnet M_2 by placing a diamagnetic body G (*fig. 2*) closely below it. The latter exercises an upward force of repulsion upon M_2. Under suitable conditions (e.g. if M_2 is sufficiently small compared to M_1) any slight lowering of M_1 from the state of equilibrium results in an increase in the repulsion exercised by G that is greater than the decrease in the attraction by $M_2 : M_1$ hence returns to the position of equilibrium. A slight upward displacement of M_1 is compensated in an analogous way. As regards horizontal displacements of M_2, the equilibrium is stable in the absence of the diamagnetic body, and this situation is not altered by introducing it, provided that its horizontal extent is sufficiently large.

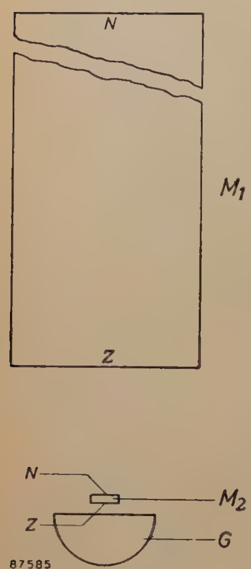


Fig. 2

We made an experimental model of the arrangement described here with M_1 a "Ticonal" bar magnet 30 cm long and 3 cm in diameter, and a body G of very pure graphite, for which $\mu_r = 0.9999$ (about the strongest diamagnetism attainable in solids, at normal temperatures.) The levitated magnet M_2 was a cylindrical ferroxdure disc with a diameter of 1 mm and a thickness of 0.3 mm.

Other arrangements for producing levitation are possible when diamagnetic substances are used. Braunbek succeeded in levitating a piece of graphite weighing 75 milligrams between the specially shaped poles of an electromagnet, with a field strength between the poles of about 1.8×10^6 A/m (23 000 oersteds). This extremely great field strength was required because the levitation is based on the homogeneity of the field, which must be greater according as μ_r approaches unity and which has to extend over a distance comparable to the dimensions of the levitated body.

A field H exerts on a volume element $\mathrm{d}V$ of a substance with $\mu_r < 1$ a force in the direction z given by

$$-\tfrac{1}{2}\mu_0 (1-\mu_r) \frac{\partial H^2}{\partial z} \mathrm{d}V$$

where μ_0 is the permeability of free space. If this force is to compensate the weight $\varrho g \mathrm{d}V$ (ϱ being the density, g the constant of gravitation), then we must have

$$\frac{\partial H^2}{\partial z} = -\frac{2g}{\mu_0}\frac{\varrho}{1-\mu_r}. \qquad (1)$$

For graphite, $\varrho/(1-\mu_r)$ is 2.3×10^7 kg/m³; for most other diamagnetic substances it is far greater. For the levitation of graphite the field must therefore obey the relation

$$\frac{\partial H^2}{\partial z} = 3.6 \times 10^{14}\ \mathrm{A^2/m^3}\ (\approx 5.7 \times 10^8\ \mathrm{Oe^2/cm}).$$

The very large field strength required by Braunbek could be obtained only with the aid of an electromagnet, so that he did in fact require energy for this levitation. The power required was very great, viz. 7 MW/kg. A dimensional analysis of the problem shows that for a given configuration, the maximum value of H required to obtain a given inhomogeneity ($\partial H^2/\partial z$) decreases with the square root of the dimensions. In this way it proved possible to repeat Braunbek's experiments in a smaller scale, using a *permanent* magnet, as shown in *fig. 3*.

Apart from diamagnetic materials there is another group of substances with $\mu_r < 1$, viz. superconductors, for which $\mu_r = 0$. A far smaller degree of inhomogeneity of the levitating field is therefore adequate (cf. eq. 1), so that it would be possible in an arrangement similar to that of fig. 3 to levitate a comparatively large superconducting body by means of a far weaker magnet.

Using superconductors, the latter arrangement may be reversed to accomplish a spectacular result: a permanent magnet can be floated freely above a cup-shaped superconducting body [7]). This experiment may be elucidated as follows. Consider first the case of two ferroxdure discs, one placed above the other, magnetized in the direction of their axes and repelling each other (*fig. 4a*). If the lower disc is held steady at a certain height, the top one is in a state of stable equilibrium as regards vertical displacement; however it can never at the same time be stable for horizontal displacement as well (Earnshaw's theorem). Now let us imagine the lower magnetic disc being removed and the space below the former plane of symmetry being occupied by a superconductor (fig. 4*b*). This will act upon the remaining magnetized disc with exactly the same repulsive force as before, since the magnetic field

[6]) W. Braunbek, Z. Physik **112**, 753-763, 1939, also **112**, 764-769, 1939.

[7]) The experiment, in which a magnet weighing about 1 g was used and cup and magnet were placed in a Dewar flask above liquid helium, has been described by V. Arkadiev, Nature **160**, 2, 330, 1947. The cup shape is used to give stable instead of a neutral equilibrium in the horizontal direction.

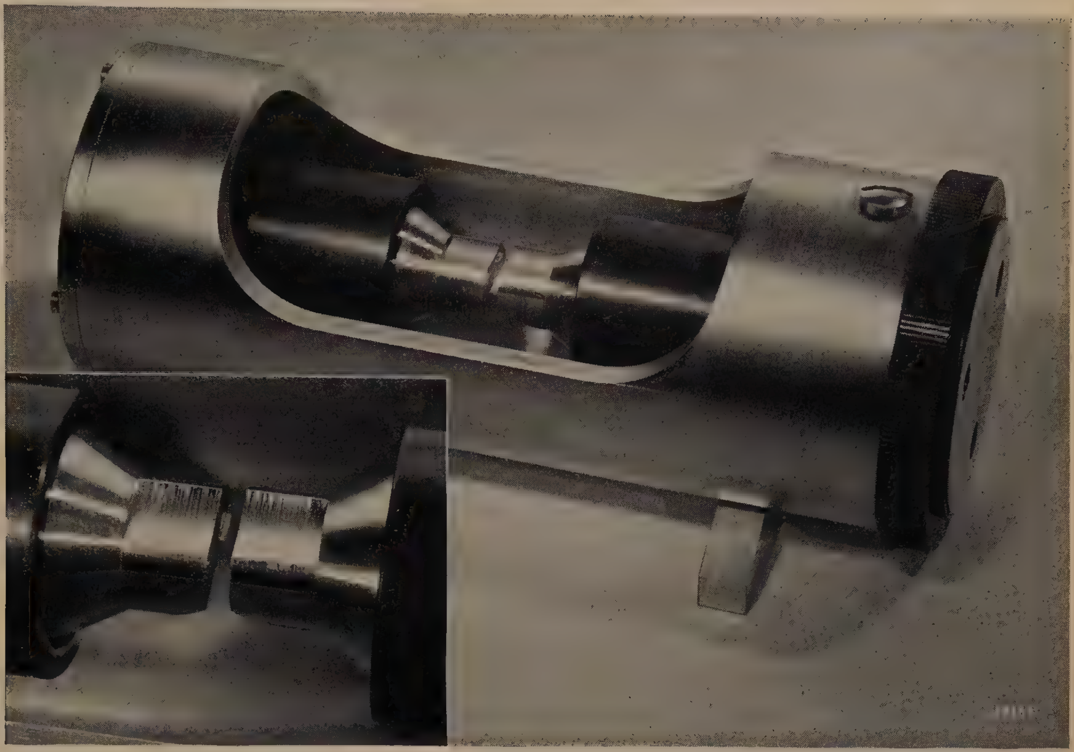


Fig. 3.

above the superconductor can be expressed as the sum of the field of the remaining magnet and that of its *mirror-image* in the surface of the superconductor [8]). The magnet thus remains poised above the superconductor, but the difference from the case of fig. 4*a* is that now the mirror image remains at all times vertically below the magnet, thus obviating the instability of the equilibrium for horizontal displacement.

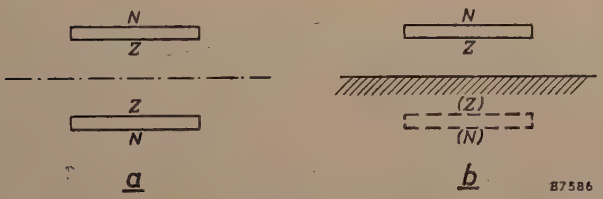


Fig. 4.

If we finally imagine a diamagnetic substance being substituted for the superconductor, the same argument remains valid; only the magnetic strength of the mirror image will be reduced by a factor of $(1 + \mu_r)/(1 - \mu_r)$, which means by a factor of 2×10^4 for graphite. Simple dimensional analysis demonstrates, however, that this far weaker repelling force will still be adequate to compensate the weight of the magnet, if only the latter is small enough. A permanent-magnet particle of suitable shape and with dimensions of a few microns should be capable of floating freely at a very low height, likewise of the order of a few microns, above the surface of a diamagnetic substance such as graphite.

A. H. BOERDIJK.

[8]) For completeness, it should be pointed out that substitution of a substance with $\mu_r \doteq \infty$ (ideal ferromagnetic) for the superconductor would be equivalent to creating a mirror image of the object magnet with *reversed* polarity, i.e. with its magnetization in the same direction as that of the object magnet.

COLD-CATHODE TRIGGER TUBES

by C. H. TOSSWILL *).

621.373.444:621.387

The triode cold-cathode glow-discharge tube made its first appearance about twenty years ago as an alternative to the hot-cathode thyratron: it consumed no stand-by power and was immune to heater or filament breakage, but could only be used where the current demand was small and where the high maintaining potential and slow decay of the cold-cathode discharge were acceptable. These triodes have found many applications, notably as selectors in remote party-line telephone-switching; as counting elements in rings and other arrays; and as information storage devices, where their irreversible triggering action has been exploited. This article describes some of the improvements made in recent years in the design of cold-cathode trigger tubes.

Introduction

The work to be described here began in response to certain criticisms of existing cold-cathode tubes. It has had two distinct stages: first, the design of a general purpose experimental cold-cathode tube with more closely-defined and stable characteristics, and second, the development of two specialised forms suited to service in radiation monitors [1]).

Principle of operation

The triode cold-cathode tube is founded upon two fundamental properties of the low-pressure glow-discharge: first, the potential difference across a gap needed to initiate a glow-discharge is greater than that subsequently required to maintain it, and, second, the initiating or breakdown potential can be reduced by the insertion of a subsidiary or trigger discharge. Therefore, if a potential lying between the breakdown (V_b) and maintaining (V_m) values is applied to the main gap, no discharge will pass until a suitable trigger discharge is supplied. The main discharge will then continue until the voltage across the main gap is reduced to below the maintaining potential V_m. As the current carried by the main discharge may greatly exceed that carried in the trigger discharge, a power gain is available.

In a triode, the trigger discharge passes between one main electrode and a third, "trigger", electrode, and has an effect dependent upon its duration and the tube geometry. The relation between minimum steady trigger current (I_t) and main gap or anode potential (V_a) shown in *fig. 1* applies to the case where both discharges have a common cathode, and is often known as the *transfer characteristic*. Certain time delays, the statistical and formative lags (see below), affect the development of the trigger discharge itself. Always, however, the influence of the trigger is restricted to the beginning of the main discharge; after this, the trigger has no effect and the only means of interrupting the discharge is to lower the main gap potential beneath V_m. Such reduction causes the gas to revert to the condition prior to the passage of the discharge, to be "de-ionized", in a certain characteristic time. In the method of extinction and in the irreversible action of the control electrode the cold-cathode triode thus resembles the hot-cathode thyratron.

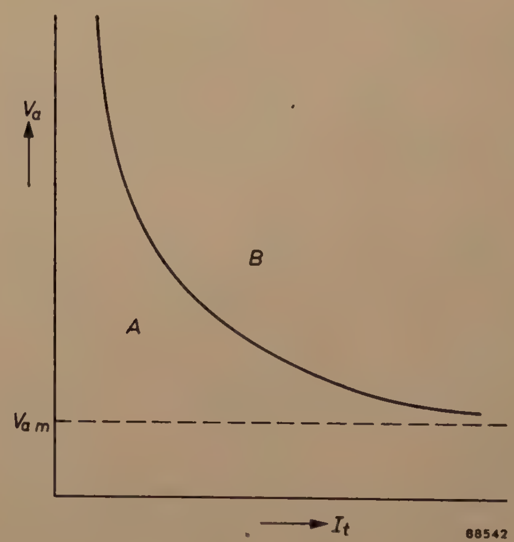


Fig. 1. Transfer characteristic of a trigger tube. This curve gives the minimum steady trigger current I_t necessary at a given anode voltage V_a to initiate breakdown. Once breakdown occurs (at the line dividing regions *A* and *B*) the tube continues to conduct so long as the anode voltage remains above the maintaining potential V_{am}.

The breakdown characteristics of cold-cathode trigger tubes may be represented by a closed loop

*) The work described in this article was carried out at Mullard Radio Valve Co., Mitcham, Surrey.

1) E. Franklin and J. Hardwick, The design of portable gamma and beta radiation measuring instruments, J. Brit. Inst. Radio Engrs. **11**, 417-434, 1951. D. Taylor, Radiac instrumentation, J. sci. Instr. **29**, 315-322, 1952.

such as that shown in *fig. 2* which refers to a triode, the Mullard Z 300T (*fig. 3*). Points lying inside this loop correspond to anode and trigger voltages which do not cause a discharge. In order to fire the tube, the anode voltage and/or trigger voltage must change in such a way that the loop is crossed somewhere.

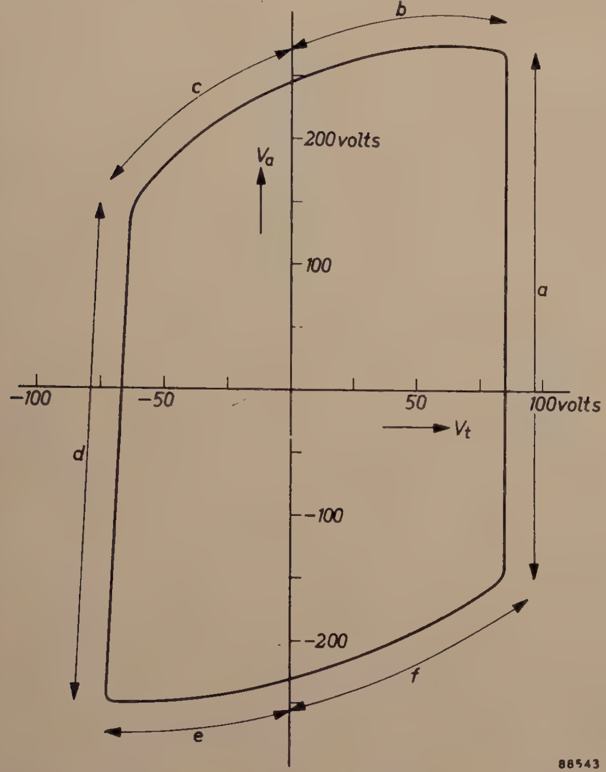


Fig. 2. Breakdown characteristic of Z 300T tube. For the tube to fire the closed loop must be crossed somewhere. The configuration of the trigger discharge depends on where the loop is crossed (regions *a-f*, see text).

Once the line has been crossed and the discharge initiated, the anode and trigger voltages may once more correspond to a point inside the loop. The section *a* of the loop refers to the firing of the tube by a discharge from trigger to cathode (direction of positive current flow). It is evident from the graph that the triggering voltage required is here independent of the anode voltage. A discharge from anode to cathode occurs when section *b* of the loop is crossed. Section *c* refers to a discharge from anode to trigger, section *d* from cathode to trigger, section *e* from cathode to anode and section *f* from trigger to anode.

The barium oxide layer on the cathode of the Z 300T tube gives it a work function lower than that of the anode, so that the tube has rectifying properties. For given supply voltage and series resistance the current flow from anode to cathode and trigger to cathode is therefore larger than in the reverse direction. For this reason, this tube is operated in quadrant *1* of fig. 2, where anode and trigger voltages are positive.

Mechanism of breakdown; statistical and formative lags

When a potential difference of a few volts is applied between cold electrodes in a gas, the latter behaves substantially as a perfect insulator [2]). Ionization due to cosmic radiation will give rise to a very small current, which will increase to a saturation value as the field is increased to a point where all the free electrons and ions are drawn to the electrodes. This saturation current is not constant with time, however, being subject to wide statistical variations owing to the random nature of the ionization. The saturation current will be greater and largely free of fluctuations if the cathode is exposed to light so as to emit a copious supply of photo-electrons.

As the applied field is increased, the free electrons become capable of imparting energy to the outer electrons of the gas atoms: the latter may be excited in the normal way and emit radiation, or they may be raised to metastable states, or, when the colliding electron has sufficient energy, some atoms may be ionized. With the gas pressures and electrode separation commonly used in trigger tubes, the mean free path of the free electrons will be only a fraction of the inter-electrode distance: hence ionization by collision will occur only when the applied potential is several times the ionization potential of the gas.

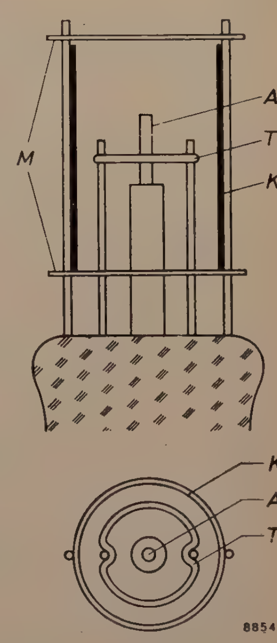


Fig. 3. Electrode structure (schematic) of the Z 300T cold-cathode tube. *A* anode, *T* trigger electrode, *K* cathode, *M* mica spacers. The tube is filled with argon to a pressure of 8 mm Hg. The electrodes are of nickel, the cathode having a coating of barium oxide upon its inner face.

The positive ions in the gas acquire much the same energy as the free electrons but, because their mass is equal to that of the gas atoms, they give up on an average about half their kinetic energy at each collision. Since, also, elastic collisions are more probable than in the electron

[2]) For a general review of the field of gas discharge phenomena see F. M. Penning and M. J. Druyvesteyn, Rev. mod. Phys. **12**, 1940.

case and charge-exchange may occur, the positive ions contribute little to ionization by collision.

With the onset of ionization due to electron collision, the anode current increases rapidly with the applied potential. If the number of new free electrons created by a single electron moving through unit potential is η (the gas ionization coefficient), the number of new free electrons created by n electrons traversing an element dl between the electrodes is

$$dn = n\eta E \, dl.$$

Assuming that n_0 electrons leave the region of the cathode per second and that the field E is uniform, the number of electrons arriving at the anode per second is:

$$n_a = n_0 e^{\eta EL} = n_0 e^{\eta V},$$

where L is the electrode separation and V the applied potential.

The anode current may then be written

$$I = I_0 e^{\eta V}, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (1)$$

I_0 being the cathode emission current corresponding to some electron-supplying mechanism, e.g. photo-emission.

So long as the anode current is determined by (1) the discharge cannot become self-sustaining: if I_0 ceases, (by shutting off the light in the case of photo-emission) the anode current will vanish. Physically, this means that when the free electrons initially present and those created by ionization have been drawn to the anode, the anode current must cease unless further free electrons are supplied. Hence, for the development of a self-sustaining discharge a secondary electron-supplying process is required. This may be effected by the positive ions, which on striking the cathode, give rise to further electron emission which enables the discharge to become self-sustaining. The number of new cathode electrons emitted per ion-pair formed between the electrodes is termed the secondary ionization coefficient γ. It is closely related to the work function φ of the cathode; for a given gas, γ is larger the smaller φ.

Taking this secondary process into account, the relation between I and I_0 becomes:

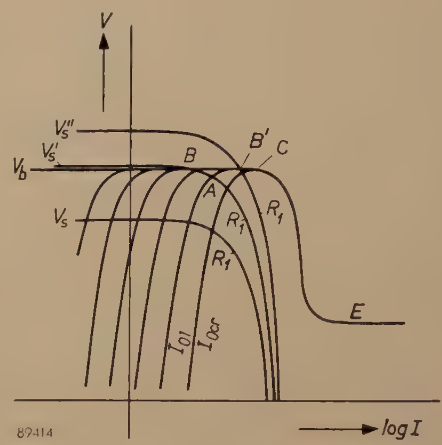
$$I = I_0 \frac{e^{\eta V}}{1 - \gamma[e^{\eta V} - 1]}. \quad . \; . \; . \quad (2)$$

The characteristics respresented by this expression are plotted in fig. 4 for various values of the initial cathode current I_0.

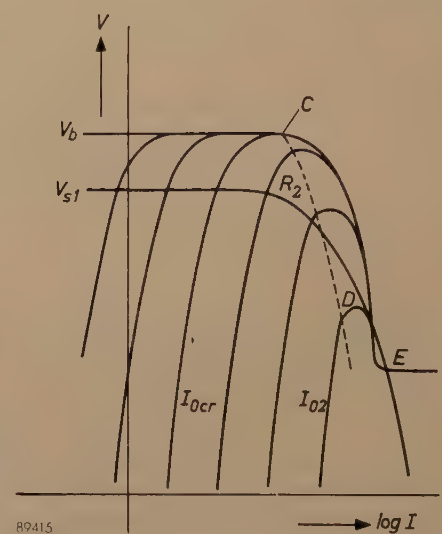
When the voltage V across the electrodes has the value V_b defined by

$$\gamma[e^{\eta V_b} - 1] = 1, \quad . \; . \; . \; . \; . \quad (3)$$

the conclusion from (1) that $I \to 0$ as $I_0 \to 0$ is no longer necessarily valid, for equation (2) then leaves I indeterminate. What actually happens can be seen from *fig. 4*. The voltage V across the tube and current I are determined for a given value of I_0 by the intersection of the relevant curve with the load line R (curved owing to the logarithmic scale)



a



b

Fig. 4. Voltage-current characteristics (schematic) for a discharge between two electrodes in a gas, for various values of the initial current I_0. For clarity the current is plotted logarithmically. The numerical value of I_0 for each curve is given by its intercept on the current axis.
a) Normal self-breakdown, with $I_0 < I_{0cr}$. This refers to breakdown in the trigger gap. R_1 are load lines (curved owing to log scale) corresponding to supply voltages V_s, V_s' and V_s''. When the supply voltage exceeds V_b (e.g. V_s') breakdown occurs and the current which then flows for an initial current of I_{01} (say) is given by the point A. If I_0 then $\to 0$, the discharge stabilizes at B. For a higher supply voltage (V_s'') the discharge stabilizes at B'. The trigger discharge may also burn in the glow-discharge region E if the series resistor R_1 is small enough.
b) Triggered breakdown ($I_0 > I_{0cr}$) of the anode gap. When $I_0 > I_{0cr}$, space charge effects begin to appear which cause the breakdown voltage to be lowered. The dotted curve from the point C represents the locus of the points at which space charge begins to have an effect. For an initial current of I_{02}, the breakdown voltage is lowered to the point D. R_2 is the load line corresponding to a supply voltage V_{s1} and series resistance R_2 and is tangential to the curve I_{02}. In the absence of I_{02} the tube will be non-conducting but when I_0 is raised to the value I_{02} by the trigger discharge, breakdown occurs and the tube current stabilizes at the point E.

corresponding to the supply voltage V_s and the resistance in series with the source.

As the supply voltage V_s is raised (fig. 4*a*), the load line is displaced parallel to the ordinate and at a voltage somewhat higher than V_b a self-sustaining discharge becomes possible; the voltage and current between the electrodes are then represented by a point such as *A* (fig. 4*a*). The discharge is self-sustaining for, as $I_0 \rightarrow 0$ the working point on the load line moves to the left until intersection occurs at a voltage $V = V_b$ (e.g. point *B* in fig. 4*a*).

With further increase in V_s the current I increases ($B \rightarrow B'$) but V remains at V_b. At a certain value of I (point C) space charge effects begin to distort the hitherto uniform field; further increase in I is accompanied by a fall in the voltage V across the electrodes and by the development of the glow discharge (region E in fig. 4*a*).

For values of the initial cathode current I_0 less than a certain critical value I_{0cr}, the breakdown voltage (i.e. the maximum voltage V_{max} reached in the establishment of a self-sustaining discharge) has a constant value V_b determined solely by the materials, geometry and gas filling of the tube. Above this critical value, however, the breakdown voltage V_{max} becomes a function of I_0. Space charge effects now appear at an early stage and prevent V from reaching V_b. Consider the characteristic I_{02} ($> I_{0cr}$) and a series resistor R_2 giving the load line R_2 in fig. 4*b*. As the supply voltage V_s is increased to the value V_{s1}, V passes its maximum value near D and breakdown occurs. The voltage V across the electrodes drops sharply and, for a not too large value of the load (i.e. $\leqslant R_2$ in this case), current and voltage stabilize at a point such as E corresponding to the glow discharge regime.

Normal *self-breakdown* between two electrodes in a gas (fig. 4*a*) is the process taking place in the *trigger* gap of a cold-cathode tube. I_0 is normally much less than I_{0cr}. A pulse with a voltage greater than that corresponding to the gap breakdown voltage V_b (hereafter designated V_{tb}) causes a self-sustaining discharge to be set up in the trigger gap for the duration of the trigger pulse. (A condenser is often connected between trigger and cathode to intensify or prolong the trigger discharge.) A fraction of the charge carriers of this discharge are transported to the main anode-cathode gap and form the "I_0" to initiate the main discharge, as outlined below. (Whether the trigger discharge burns in the Townsend region — the current region up to the point C, fig. 4 — or in the glow discharge region near E, depends on the circuit conditions. These must, of course, be so chosen that the trigger discharge is of sufficient intensity to provide the "I_0" necessary to ignite the main gap — see below.)

Triggered breakdown is the process by which the *main anode-cathode* discharge is started. The voltage at which a self-sustaining discharge across this gap can start is temporarily lowered by providing an initial current I_0 that exceeds the value of I_{0cr} for this gap (fig. 4*b*). As mentioned above, this I_0 is derived from the already-established trigger discharge, which can easily be made to supply charge carriers at such a rate that I_{0cr} is exceeded. (The lowering of anode breakdown voltage as a function of trigger current I_t was shown schematically in fig. 1). If, then, across the main gap, a steady potential is applied somewhat below the normal self-breakdown value V_b (hereafter designated V_{ab}), the tube will remain quiescent until triggered by the presence of an adequate initial current I_0 due to a discharge set up in the trigger gap.

There is always a certain delay after the application of the trigger pulse before the self-sustaining discharge is set up in the trigger gap. The delay is made up of two components. The first, which is by no means constant, is known as the *statistical lag* and arises from the necessity for the presence of at least one favourably-placed electron to enable the discharge to start. In practice, if $I_0 \ll I_{0cr}$ an uncertain and frequently lengthy delay occurs. With increase in I_0 the statistical lag is gradually eliminated. It is desirable that the initial current I_0 lies between the value just able to remove the statistical lag and the value I_{0cr}. The inter-electrode current which then flows is called a *priming current*. With many cathodes a suitable level of ambient illumination will give rise to such a priming current; other means for providing a priming current will be described below.

The other delay is the time interval required to build up the self-sustaining discharge from the moment the first favourably-placed electron appears until the discharge has reached its equilibrium value. This delay is called the *formative lag;* it does not vary randomly but it depends sharply on the excess voltage of the trigger pulse above the breakdown voltage V_b of the trigger gap. To keep this delay small, the trigger pulse voltage should exceed the trigger breakdown voltage by some tens of volts (e.g. V_s'' in fig. 4a).

Desirable characteristics of cold-cathode tubes

To widen the field of application of trigger tubes the following should be aimed at:

a) Good reproducibility of breakdown and maintaining potentials, both from tube to tube and in individual tubes during life.

b) Faster operation, if possible approaching the response of hot-cathode devices.
c) Improvement of input sensitivity.
d) Reduction or elimination of random operating delays, even at low levels of ambient illumination.
e) Reduction of incremental impedance and noise level; for some uses, e.g. for speech-channel applications, the noise level of most tubes is too high.
f) Small dimensions.

It would not be possible to improve on all these points simultaneously, as some of them are partly contradictory. The initial aim, therefore, was to evolve a design and a manufacturing technique to give a cold-cathode tube in which some of the existing characteristics were improved, viz. points *a*, *d* and *f*: uniformity of characteristics and stability during life, freedom from random delays (statistical lags) and robust miniature construction. These particular properties were selected for improvement because it appeared that they were not incompatible and also because they are collectively of importance in many applications [3]).

It will be noted that all three features are ingredients of reliability, the improvement of which was indeed the primary aim. It will further be noted that while the raising of the intrinsic input sensitivity was not considered at this stage, the effective sensitivity of the whole circuit is of course improved when the tube characteristics are accurately known and maintained, for the input signals may then safely be smaller without fear of the tube not responding or, worse, spurious breakdown when not signalled.

Improvement of reliability and performance

The breakdown and maintaining potentials of cold-cathode tubes are inversely related to the values of the gas ionization coefficient η and the secondary ionization coefficient γ. It seems obvious therefore to choose gas fillings and tube materials which yield large and stable values of both coefficients: in fact, the inert argon filling and barium-on-nickel cathode of the Z 300T tube were selected on this basis. However, the fragile nature of the barium layer (which is susceptible to change even with normal discharges and is removed entirely by sputtering if temporarily overloaded) together with gaseous contaminants from the wall of the tube and elsewhere, lead to changes in η and γ which are observed by the user as the objectionable variations mentioned above. The steps taken to stabilize these quantities are outlined below.

Another matter requiring attention is the elimination of the statistical lag. As we have seen, to do this, a priming current must flow in the otherwise dormant tube.

Molybdenum sputtering technique

In 1946 Penning and others [4]) described means of stabilizing the basic glow discharge quantities for certain materials prior to the successful measurement of the "normal cathode fall" (the constant drop in potential near the cathode surface, which occurs when the current through the tube is between certain limits). The sustained cathode sputtering technique employed in this work had obvious application in any glow-discharge device dependent upon stability of the operating potentials, and was soon employed in the voltage reference tubes 85A1 and 85A2, where a molybdenum cathode is used in conjunction with a filling of neon and argon [5]). Similar arguments clearly favoured the introduction of this system in the class of trigger-tube under consideration. There are the following objections to the molybdenum-sputtering technique. Working potentials are high compared with those of tubes using barium or, say, potassium cathodes. Photo-emission arising from normal ambient illumination is ineffective in supplying the priming current necessary for the elimination of the statistical lag. Lastly, the manufacturing technique is not without its difficulties and the price of the materials is a factor of some significance. However, the advantages greatly outweigh the objections.

The successful application of the cathode sputtering technique to trigger-tube manufacture demands an electrode structure quite different to that of Z 300T (fig. 3). Thus, in the first place, mica spacers must be abandoned. They act as barriers to the transport of sputtered material to the walls of the tube, and when coated with this material, act as a source of inter-electrode leakage. Also, mica is a laminar material, and often secretes contaminants which can be slowly released during the operation of the tube. Secondly, the cylindrical form of the

[3]) Others have aimed at improvements in directions not attempted in the present work. An experimental tube suitable for speech channel applications, the Z 500T, has been developed in the Philips Laboratories in Eindhoven; see J. Domburg and W. Six, Philips tech. Rev. 15, 265-280, 1953/54. Other work in this field is reported by M. A. Townsend and W. A. Depp, Bell. Syst. tech. J. 32, 1371-1391, 1953. Work on improving the speed of response is described by G. H. Hough and D. S. Ridler, Electronic Eng. 24, 152-157, 1952.

[4]) F. M. Penning and J. H. A. Moubis, Philips Res. Rep. 1, 119-128, 1946; T. Jurriaanse, F. M. Penning and J. H. A. Moubis, Philips Res. Rep. 1, 225-230, 1946; and T. Jurriaanse, Philips Res. Rep. 1, 407-418, 1946.

[5]) T. Jurriaanse, Philips tech. Rev. 8, 272, 1946.

cathode is not suited to sputtering. The cathode must have a shape which allows the unimpeded escape of sputtered material, and must be so placed that while the walls are covered with a uniform film, the lead-out area remains bare. Thirdly, the tube envelope must be small enough to be covered in a reasonable time, and also be free from re-entrant portions such as exist in the old "pinch" construction.

Experiments showed that a small oblong cathode plate mounted with its plane parallel to the axis of the tube, in a standard 7-pin or 9-pin miniature envelope, met all these needs very well. This structure is shown in *fig. 5*. Subsequent field experience has shown that this rather frail-looking structure is well able to withstand the shocks of normal service.

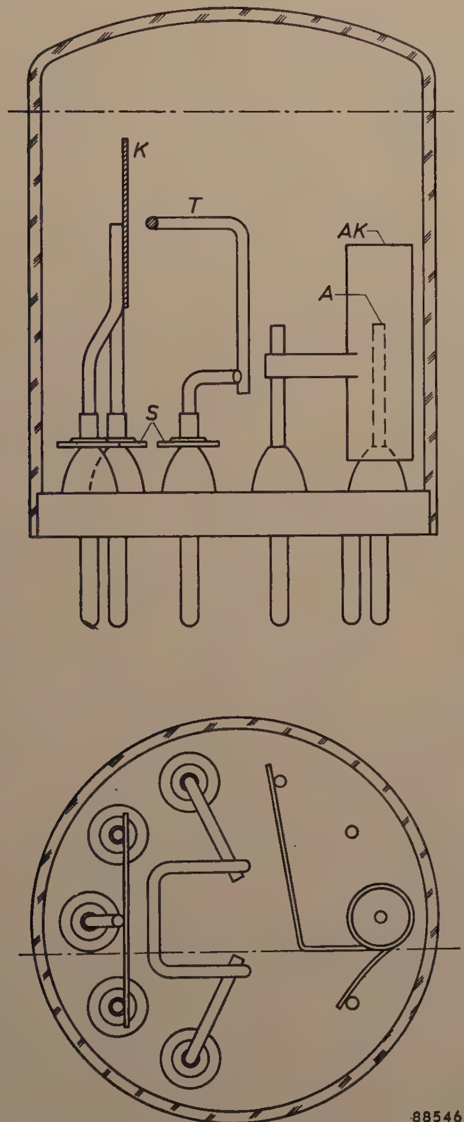


Fig. 5. Electrode structure (simplified) of experimental Mo-sputtered tetrode trigger tube. *K* cathode, *T* trigger electrode, *AK* tubular auxiliary cathode surrounding the anode *A*. *SS* are shields which leave a region near the lead-out wires free of sputtered material (see below).

Provision of priming current

Methods of providing a priming current include: *a*) The admission of normal daylight — the simplest solution, and that used in Z 300T. However, this is ineffective with the molybdenum cathode, whose photo-emission is negligible for the spectral range transmitted by normal glasses. Moreover, reliance on the photo-electric effect imposes undesired restrictions upon the conditions of use of the tube. *b*) The use of radio-active materials. This would have been a more satisfactory solution, but was rejected because of lack of experience with "soft" emitters whose influence would not be objectionable outside the tube, and because the third system has additional advantages quite apart from the priming. *c*) The use of an auxiliary discharge. Such a discharge may be direct, passing the trigger and cathode, or indirect, achieving its effect by the transport of some of its ions into the trigger-cathode space, possibly over a considerable distance.

The direct auxiliary discharge imposes certain restrictions on the trigger circuit, but is adopted in one of the tubes described below (Z 801 U). An upper limit somewhat below 1 micro-ampere is set on the permissible value of such a direct priming current by the need to avoid distorting the applied fields within the tube by the accumulation of space-charge.

The indirect discharge, which may be larger, is used in the stabilizer tube Z 800 U — with the discharge flowing to the anode from an auxiliary cathode. The latter is of tubular form, and surrounds the wire anode (fig. 5). This tubular auxiliary cathode provides an economical solution to two problems: how to obtain within a miniature envelope an effective spacing large enough to create a useful difference between V_{ab} and V_{am}, and how to screen a discharge between anode and auxiliary cathode from the trigger-cathode space. By varying the recession of the anode tip within the auxiliary cathode, V_{ab} may be altered without affecting the remainder of the trigger-tube, while the presence of the auxiliary discharge stabilizes the potential of the auxiliary cathode relative to the anode. The auxiliary discharge may now reach 10 μA, and so enter the current range controllable by standard resistors, without exerting an excessive influence on the field in the trigger-cathode gap.

With the introduction of the auxiliary discharge, the maximum operating delay is reduced from about 1 sec to the order of 100 μ sec, which is essentially the formative lag.

Experimental trigger tube

The experimental tube (fig. 5) incorporating the features outlined above showed satisfactory behavi-

our. In order to keep the triggering voltage V_{tb} to a minimum, the trigger-cathode separation was kept small (1-2 mm). In addition, a trace of argon (about 1%) was added to the gas filling of neon, at a pressure of 40 mm Hg. Under these conditions $V_{tb} \approx$ 115 V.

The recession of the anode tip within the tubular auxiliary cathode is effectively limited by the fact that, however much the recession, there is a limit to the attainable anode-cathode breakdown voltage (V_{ab}) fixed by the breakdown between auxiliary cathode and cathode. Beyond a certain point, therefore, further recession will not enhance the value of V_{ab}. For the gas filling used, V_{ab} has its maximum at 250 V. Under these conditions, the discharge-maintaining voltage V_{am} = 100 V.

The triggering voltage stability of early experimental tubes was not entirely up to expectation. This was found to be due to contamination of the trigger electrode, which was then of nickel. So long as different materials were used for trigger and cathode, it proved impracticable to clean both by sputtering. The use of a molybdenum trigger gives a satisfactory stability, but it has the disadvantage that in circuits in which the trigger may become earthed, the maximum operating voltage may have to be smaller to preclude spurious trigger-anode breakdown.

Variations in the discharge-maintaining potential V_{am} are similar to those of the voltage reference tube (diode) 85A2 and are satisfactory.

Self-quenching operation of trigger tubes

The achievement of better stability and the virtual elimination of statistical lags cleared the way for the design of two trigger tubes for certain specific applications, viz. for use in a stabilizer circuit (Z 800U) and for use in a rate meter circuit (Z 801U) [6]). The former has to respond to very small continuous currents and the latter to very small pulses of current.

While both the Z 800U and the Z 801U may be used in certain conventional cold-cathode tube circuits, they are primarily designed for the oscillatory self-quenching circuit. In this type of circuit (*fig. 6*), the anode current is derived from the discharge of a condenser and is therefore intermittent, passing as a single pulse each time the tube is triggered. The self-quenching circuit has the advantage that the trigger electrode is capable of exercising a quasi-continuous and reversible control over the mean anode current. This gives the circuit a number of important applications. Trigger tubes working in such circuits face conditions differing markedly from those prevailing in steady current circuits. Of particular importance are the factors governing extinction of the discharge after each current pulse. The self-quenching action is obtained with an anode load R_a which is considerably less than might be suggested from the fact that to extinguish an

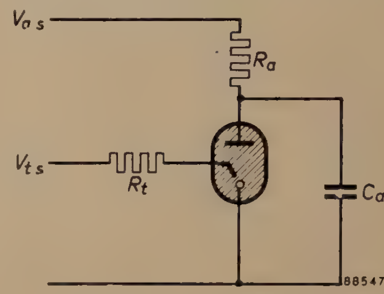


Fig. 6. Trigger tube in self-quenching circuit.

established *steady* discharge (the supply voltage remaining constant) R_a would have to be increased to at least $10^8\ \Omega$. The self-quenching circuit can operate with R_a only a hundredth of this value.

A study of the self-quenching discharge leads to the following explanation. The absence of a resistor limiting the condenser discharge enables anode breakdown to be followed by an exceptionally large tube current; this has passed its peak and largely decayed by the time that C_a has discharged sufficiently for V_a to fall to the normal maintaining voltage V_{am}, and has, in consequence, created an abnormal positive ion cloud around the anode. As this cloud moves away, the anode is able to descend well below V_{am} in an "overswing"; at the end of the overswing the flow of current through R_a carries V_a up again to $V_{a,m}$, but not before the process of de-ionization has gone far enough to lead to the collapse of the discharge. The overswing, which gives the tube the appearance of having inductive properties, commonly carries V_a down to $^3/_4$ of V_{am}. In the case of the Z 800U tube other considerations led to a design having satisfactory self-quenching properties; with the Z 801U, however, the main problem was the reconciliation of the self-quenching requirements with the remainder of the specification.

Stabilizer trigger tube

The trigger tube Z 800U [7]) (*fig. 7*) was originally developed for use in a stabilized H.T. supply for Geiger counters in radiation monitors [1]), viz. the low-current stabilizer circuit introduced by Goulding [8]) (*fig. 8*). The tube has to meet the following three conditions:

1) V_{ab} to be approximately 300 V.
2) The tube to operate in a self-quenching circuit.
3) The trigger input current demand to be less than 10^{-7} A.

[6]) These two special-purpose tubes are not available commercially from the Philips Electronic Tube Division.

[7]) C. H. Tosswill, British Patent No. 709 103.
[8]) F. S. Goulding, British Patent No. 732 776; and, A variable voltage stabilizer employing a cold-cathode triode, Electronic Eng. 24, 493-497, 1952 (the Z 800U is referred to as VX 8107, and the Z 801U as VX 8086); G. O. Crowther, Cold-cathode voltage stabilizer, Electronic Eng. 25, 127, 1953.

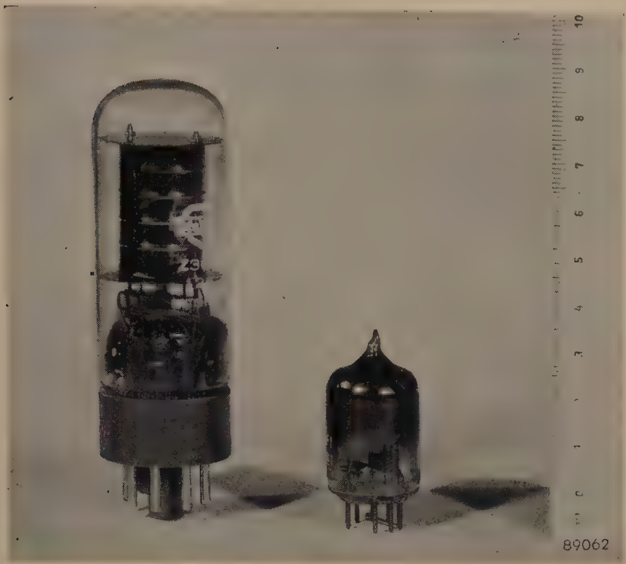


Fig. 7. Photograph of the Z 800 U trigger tube (centre, see also fig. 5) showing its size in relation to the Z 300 T tube mentioned earlier (fig. 3). (The scale on the right is in cm.) The tubular auxiliary cathode and the sputtering shields round the lead-in wires can just be seen through the envelope of the Z 800 U. The ratemeter trigger tube Z 801 U described later (see fig. 14) and the general purpose tube Z 803 U are identical in size and appearence.

With the same electrode structure as that used in the experimental tube (fig. 5), the required breakdown voltage was obtained by using a gas filling of neon with 3% argon at a pressure of about 40 mm Hg. This percentage of argon is somewhat greater than that used in the experimental tube (1%). The change also raises the maintaining voltage V_{am} to 105 V and the triggering voltage V_{tb} to 125 V.

The tube has to operate under circuit conditions which create self-quenching discharges in both the trigger-cathode and the anode-cathode gaps (*fig. 9a*, in which the priming electrode is omitted for simplicity). Both anode and trigger currents therefore

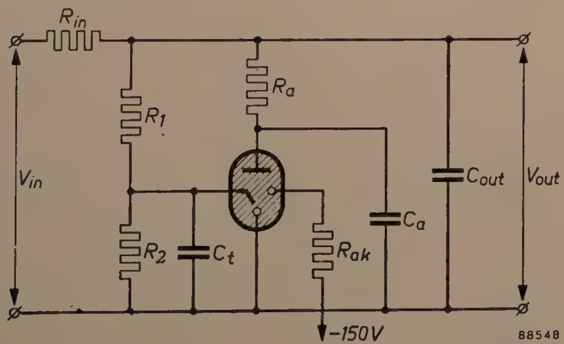


Fig. 8. Goulding's stabilizer circuit [8]) using the trigger tube Z 800U. R_1 = 68 MΩ, R_2 = 100 MΩ, C_t = 270 pF, R_a = 1 MΩ, C_a = 0.005 μF, R_{ak} = 100 MΩ, C_{out} = 0.25 μF. In this stabilizing circuit a fraction $R_2/(R_1 + R_2)$ of V_{out} is applied to the trigger electrode. Therefore if $V_{out} > V_{tb}(R_1 + R_2)/R_2$ the tube will be triggered and the pulse train through the tube will cause C_{out} to discharge until V_{out} has fallen sufficiently to dispose of the inequality. Any further fall in V_{out} will cut off the tube current because the discharges are self-quenching. In this way a stabilizing action is secured (provided that $I_{in} > I_{out}$) with V_{out} slightly in excess of V_{tb} $(R_1 + R_2)/R_2$.

take the form of trains of pulses and the interaction between them complicates the general analysis of the problem. We may confine ourselves to the important special case when the anode supply voltage $V_{as} < V_{a,b}$ and $R_aC_a \ll R_tC_t$. Each anode pulse is then necessarily the *immediate* consequence of a trigger pulse, though individual anode waveforms are independent of the actuating trigger discharges (fig. 9*b*).

The mean anode current bears a constant relation to the mean trigger current, since both are determined by condensers charged to the (fixed) supply potential V_{as} and to the trigger breakdown potential V_{tb}. Since the anode current pulse train can be stopped as well as started by the action of the trigger, i.e. by variation of the trigger supply potential V_{ts}, the mean anode current $\bar{I}_a$ can be continuously controlled over a considerable range by the trigger supply voltage V_{ts}. The circuit of fig. 9*a* may there fore be used as a D.C. amplifier. Its amplification factor is Q_a/Q_t, i.e. the ratio of the charges on the anode and trigger capacitors.

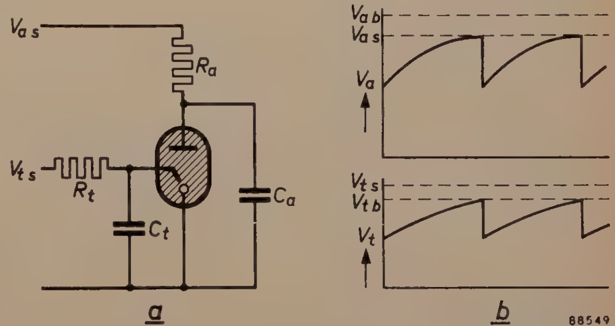


Fig. 9. *a*) Circuit giving self-quenching discharges in both anode-cathode and trigger-cathode gaps. Priming electrode (auxiliary cathode) omitted for simplicity. R_t = 100 MΩ, C_t = 470 pF, R_a = 1 MΩ and C_a = 2000 pF.
b) Anode and trigger waveforms for the circuit shown in (*a*). When $R_aC_a < R_tC_t$, a trigger pulse above a certain amplitude gives rise to an immediate anode pulse. The anode waveforms, however, are independent of the form of the actuating trigger discharge. (The breakdown and extinction processes occupy an interval that is short compared with R_aC_a and R_tC_t, and therefore appear in the graph as a vertical line.)

Operation under the above conditions implies that both discharges are self-quenching (which can be achieved without using the very high component values which would be necessary under steady operating conditions) and that the charge passing in a single trigger pulse must be sufficient to cause anode breakdown. This requirement is analogous to the I_t/V_a relationship (fig. 1) for triggering in the case of the steady current circuit. In the self-quenching case one of the variables is thus a component value (C_t); however it is more fruitful to consider another approach, viz. the effect of the trigger supply potential. (The variation of V_{ab} with C_t can be deduced from fig. 12.)

Critical trigger supply potential (V_{tsc})

Fig. 10 shows the anode supply potential V_{as} plotted against the trigger supply potential V_{ts}. The curve represents the conditions under which a flow of mean anode current $\bar{I}_a$ can just be detected. The value of V_{ts} corresponding to the nearly vertical part XY is called the critical trigger supply potential; the tube is preferably operated in this region since V_{ts} is here nearly independent of the anode supply voltage V_{as}.

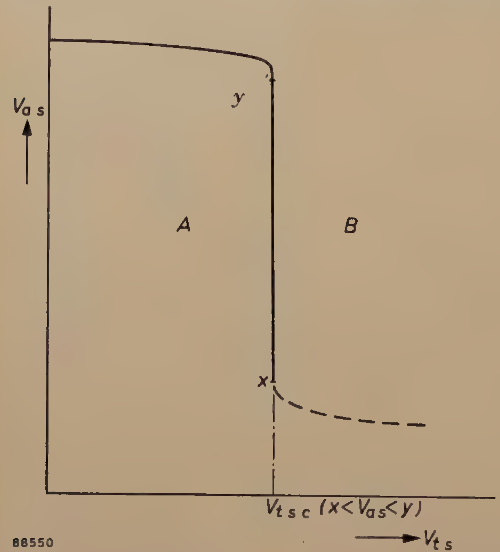


Fig. 10. V_{as} plotted against the value of V_{ts} required to just give a mean anode current, for a self-quenching circuit. Within a considerable range of anode supply voltages (X-Y) the required trigger supply potential is independent of V_{as}. This value is termed the critical trigger supply potential, V_{tsc}.

Further penetration of region B (fig.10) leads to a growth of $\bar{I}_a$; on return to region A, $\bar{I}_a$ is cut off. This provides an interesting comparison with the irreversible action of fig. 2, where the current can assume only a certain fixed value.

The critical trigger supply potential V_{tsc} can be resolved into three components:

a) The breakdown voltage V_{tb} necessary across the actual trigger-cathode space to initiate breakdown. This has been discussed in the previous sub-section.

b) The potential difference developed across R_t (fig. 9a) due to ohmic leakage over the tube surfaces. This has been reduced to negligible proportions by reducing the leakage current to 10^{-10} A. Two simple measures were sufficient to achieve this: first, the mounting of small shields upon each lead-out wire close to the point of entry into the glass, so as to create a small bare zone within which no sputtered material could fall; second the spraying of the exterior of the envelope in the neighbourhood of the lead-out wires with a silicone solution to inhibit the formation of conducting films of moisture.

c) The potential difference developed across R_t due to the pre-breakdown trigger-discharge current I_{tb} (the trigger condenser does not contribute to this current at this early stage). The maximum value of I_{tb} determines the current drain on the trigger supply which, as stated, must not exceed 10^{-7} A. I_{tb} is dependent on a number of circuit and tube parameters; these will now be discussed.

Trigger supply current drain at breakdown (I_{tb})

The influence on I_{tb} of four variables was examined as outlined below.

1) Effect of anode supply potential V_{as}. *Fig. 11* shows the effect of R_t on V_{tsc}. Since the changes in applied potentials were slow, the separation of the two curves can only be due to the voltage drop across R_t, due to surface leakage current together with I_{tb}. The leakage current (10^{-10} A) accounts for a separation of only 0.006 V for the two resistances used (10 MΩ and 68 MΩ). Hence the observed separation of approximately 1 volt must be almost entirely due to I_{tb}; the value of I_{tb} is thus $10^{-10} \times 1/0.006$, i.e. 2×10^{-8} A. It will be noted that the separation of the two curves is constant within a considerable range of anode supply voltages: this implies that I_{tb} is independent of V_{as}. Fig. 11 also shows that the critical trigger supply voltage is substantially independent of V_{as} (cf. fig. 2 for the steady discharge case).

2) Effect of auxiliary cathode current I_{ak}. Fig. 11 is based on a value of I_{ak} = 3 μA. Little change in

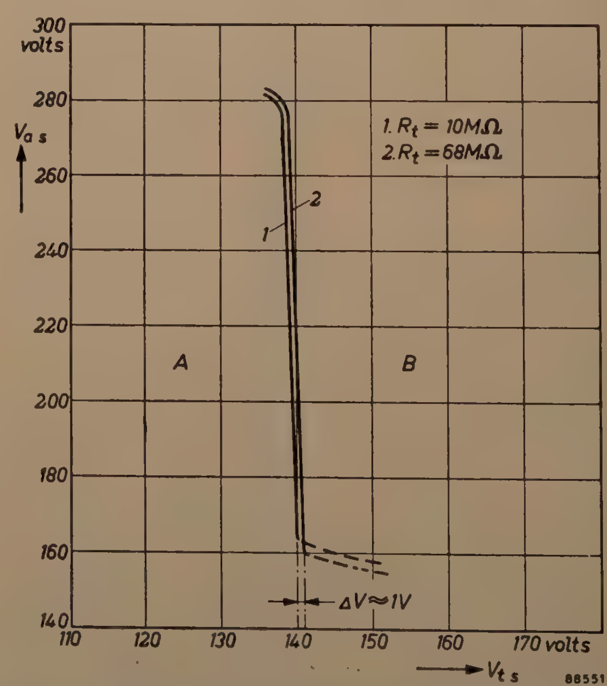


Fig. 11. Variation of V_{tsc} with V_{as} for two different values of the trigger resistor R_t. C_t = 1000 pF. From the separation of the two curves the trigger supply current drain at breakdown can be deduced (see text).

I_{tb} can be detected for variations of I_{ak} between 1 and 10 μA. This somewhat unexpected result is probably due to the fact that the auxiliary discharge starts near the end of the cylindrical auxiliary cathode and expands inward with growth.

3) Effect of trigger condenser C_t. *Fig. 12* shows again the variation of V_{tsc} with V_{as} but with R_t kept

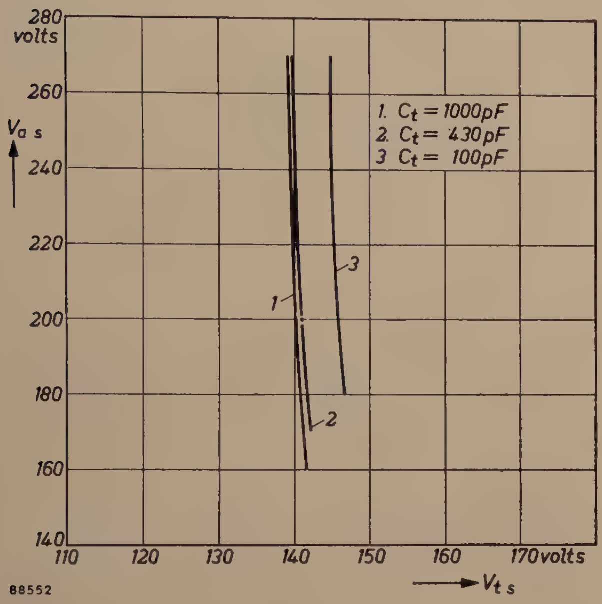


Fig. 12. Effect of trigger capacitance C_t on the V_{as} - V_{ts} characteristic. Trigger resistor R_t = 68 MΩ.

constant and three different values of C_t. The lower the value of the latter, the higher is the critical trigger supply voltage: this implies that I_{tb} is greater. Hence there is a lower limit to the capacitance of C_t, if I_{tb} is to be kept below 10^{-7} A.

4) Effect of trigger-cathode spacing. In the absence of mica spacers, a certain spread in the electrode spacings is inevitable. In tubes with smaller spacings large values of I_{tb} (10^{-6} A) were liable to occur intermittently, an effect visible as a corona surrounding the trigger wire instead of the normal glow adjacent to the cathode. This trouble was eliminated by increasing the trigger-cathode separation from 1 mm to 3 mm, whereby the trigger breakdown voltage V_{tb} is also increased, from 125 to 135 V.

Performance in stabilizer circuit

The investigations described above showed that only very minor modifications were necessary to the original experimental tube in order to get satisfactory operation in a self-quenching circuit with a very small trigger current drain.

In the restricted mode of operation defined earlier ($V_{as} < V_{ab}$, $R_aC_a \ll R_tC_t$), increase of V_{ts} above V_{tsc} gives rise to a finite mean anode current $\overline{I}_a$ whose value for a given set of supply potentials and component values depends on V_{tsc}, the anode extinction potential and the trigger extinction potential. Since V_{tsc} and the two latter quantities are constant for a given tube, the tube can be used for stabilizing a supply potential as in the Goulding circuit (fig. 8). The stabilizing action of the circuit will be clear from the figure and the details given in the caption.

This stabilizer, using the Z 800U tube, has the following important properties.

a) It works in a voltage range (200-300 V) which cannot be easily accomodated by either the glow discharge or the corona discharge diode.

b) Its current range is 5-150 μA. Diode glow discharge tubes do not normally work successfully in this range. The corona stabilizer possesses a similar current range but cannot compete in life stability with the trigger tube circuit embodying the Z 800U.

c) The output voltage can be adjusted, for each tube, by altering the resistor ratio $R_2/(R_1+R_2)$ (see fig. 8); the output voltage can in this way be varied over a considerable range to allow for the spread of Geiger-Müller tube characteristics.

The properties of the Z.800U tube may also be exploited in more conventional circuits in which the current flows continuously. Retention of the trigger condenser will still permit operation with small signal currents down to 10^{-8} A. If the auxiliary discharge be dispensed with and the occurrence of statistical lags accepted, operation with input currents down to 10^{-10} A is possible.

Rate meter trigger tube

The Z 801U [9]) was developed for use in rate meter circuits, that is to say, circuits whose output current is a measure of the rate of arrival of input pulses. The pulse source relevant to the development of this tube was a Geiger-Müller tube, but the results obtained are applicable to any system handling pulses of similar character.

The basic cold-cathode tube rate meter circuit devised by Franklin [10]) is shown in *fig. 13a*. The advantages of this circuit are described in full in the patent specification and will only be summarised here.

Each time the tube is triggered, the anode circuit goes through a single self-quenching cycle. Provided that the anode has sufficient time following the downward swing to regain the full V_{as} value before the arrival of the next input pulse, then $\overline{I}_a$ will be proportional to the rate of triggering. The connection of the trigger to the anode through a very large

[9]) C. H. Tosswill, British Patent No. 696 077.
[10]) E. Franklin, British Patent No. 638 833.

resistance has the effect of poising the trigger close to the breakdown potential. Two purposes are served in this way: a direct priming current is provided for the trigger-cathode space without the need for a fourth electrode and even the smallest input

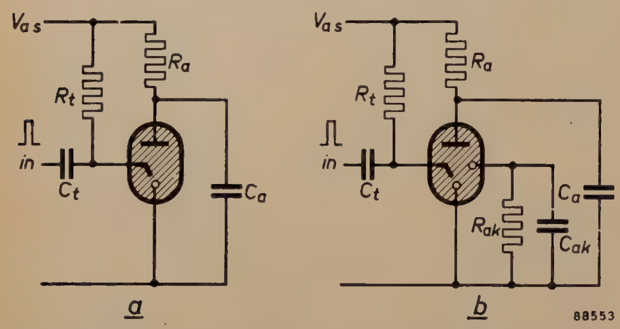


Fig. 13. *a*) Basic rate meter circuit according to Franklin [10]). *b*) Franklin's rate meter circuit using a tetrode trigger tube. Typical component values: R_t = 100 MΩ, C_t = 50 pF, R_a = 200 kΩ, C_a = 0.01 μF, R_{ak} = 1 MΩ, C_{ak} = 50 pF.

pulse applied to the trigger will cause it to exceed the breakdown potential. For the successful working of this circuit three conditions must be fulfilled:

a) The trigger priming current must not carry the trigger circuit to the point of instability, i.e. the priming current must be less than I_{tb}. Whether the trigger priming current $>$ or $<I_{tb}$ depends on the values of R_t and C_t (fig. 13*a*).

b) In this circuit, the input pulse will be able to deliver the whole of its charge (Q_{in}) into the trigger-tube. For a given value of V_{as} there is a certain minimum value of Q_{in} for anode breakdown.

c) The anode and trigger discharges must be self-quenching.

An experimental triode was constructed to study the rate meter circuit requirements. The molybdenum-sputtering technique was again used and the same neon-argon filling was used as in the stabilizer tube. The leakage-counteracting screens of the stabilizer tube — essential to keep the leakage resistances high compared with R_t ($R_t \approx 10^8$ ohms) — were also fitted. In order to meet the radiation monitor requirements, it is additionally necessary that:

a) The "extinction current" should be at least 210 μA to give an adequate value of $\bar{I}_a$ at the anticipated operating frequency. By the term "extinction current" is meant the current passing through R_a at the lower end of the anode swing, with R_a set at the minimum value compatible with self-quenching operation. The "extinction current" in the case of the triode was only 70 μA.

b) The input charge (Q_{in}) necessary for reliable response to Geiger-Müller tube signals, should be not more than about 5×10^{-11} coulomb with the anode supply voltage V_{as} safely less than the breakdown voltage V_{ab}. The value for the experimental triode was 5×10^{-10} coulomb.

c) The frequency range should preferably be about 10 kc/s. The triode had a range of only 1 kc/s.

These aims are difficult to reconcile, improvement of one being at the expense of the others. However, by considering how to increase the extinction current without loss elsewhere, we are led to the idea of separating the function of triggered breakdown from that of extinction, so as to permit of manipulating the two independently. A tetrode tube is therefore implied, though for reasons other than those obtaining in the case of the stabilizer tube. The corresponding rate meter circuit is shown in fig. 13*b*.

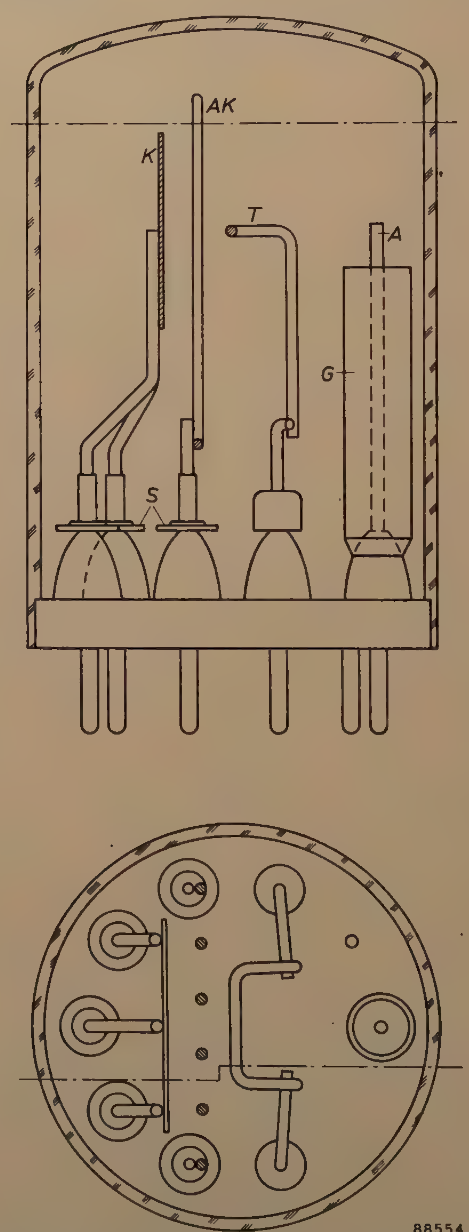


Fig. 14. Simplified diagram of electrode structure of rate meter trigger tube Z 801U. *K* cathode, *AK* auxiliary cathode, *T* trigger electrode, *A* anode, *G* glass tube, *SS* shields.

The first rate meter tetrode had as its fourth electrode ("auxiliary cathode") a perforated disc situated in the path of the discharge to the main cathode (*fig. 14*). An input signal was now followed by breakdown between anode and auxiliary cathode; then by a moderate discharge leading to a rise in auxiliary cathode potential; then by anode to main cathode breakdown; and finally by the normal downward swing of anode potential. At this point it was found, as expected, that the auxiliary cathode barrier successfully interfered with the transition to a steady glow-discharge between anode and main cathode and led to the collapse of the discharge at higher levels of extinction current than before. Trigger action remained identical with that of a triode, with the auxiliary cathode acting as cathode.

The contribution of the auxiliary cathode to extinction is that of widening the gap between the normal maintaining potential for a steady discharge and the minimum potential reached in the anode "overswing": both as a physical obstacle and as a parasitic current drain, the auxiliary cathode has a greater effect on the former potential.

The tetrode Z 801U has an extinction current of 240 μA against 70 for the triode; the input sensitivity and frequency response are unchanged. In the production version the perforated disc was changed to a wire grille. The tube has a filling of neon with 5% argon, and is shown in fig. 14.

The negative pulse circuit

The cold-cathode triode is most frequently switched by a positive signal applied to the trigger; so also were the first tetrode Z 801U tubes. The Geiger-Müller tube however is normally better used with its cathode grounded, and thus delivers a negative pulse. It was found that the tetrode could be triggered by a negative pulse [11]) applied to the auxiliary cathode (*fig. 15a*), involving a different sequence of breakdown: the input charge flows across the trigger auxiliary cathode gap; breakdown then occurs between trigger and main cathode; finally, breakdown occurs between anode and main cathode. Apart from its convenience in Geiger-Müller tube service, this negative pulse circuit has three substantial advantages. Firstly there is a remarkable gain in sensitivity: an input charge Q_{in} of 3×10^{-11} coulomb will suffice, as against the 5×10^{-10} coulomb needed with a positive pulse. Secondly, the input charge Q_{in} is constant over a wide range of V_{as}: this makes the sensitivity comparison in practice even more favourable to the negative-pulse circuit. Thirdly, since the function of the auxiliary cathode in initiating breakdown now ends with the passage of the signal charge, there is no objection to inserting an impedance in the auxiliary cathode lead which will prevent instability even if the priming current is equal to I_{tb}.

The better sensitivity of the negative-pulse circuit follows from reducing the function allotted to the input signal. The essential difference between the two modes of operation is that with the negative pulse applied to the auxiliary cathode, the trigger discharge has to be extended only over the small auxiliary-to-main cathode gap instead of over the much larger anode-to-trigger gap.

This account of the action is supported by the absence of a variation of Q_{in} with V_{as}, unlike the positive pulse circuit. Breakdown between trigger and main cathode should be largely unaffected by V_{as}; while once this stage has been passed the trigger excursion (about 40 volts) will release 4×10^{-9} coulomb, sufficient to cause anode breakdown for quite small anode supply voltages V_{as}.

Frequency response

The relation between R_a and anode-circuit time constant R_aC_a for values of auxiliary-cathode resistor between 4 and 20 MΩ is such that, for the normal anode swing of 70 volts, an adequate extinction current is obtained even at R_aC_a = 300 μsec. Against this, the trigger-circuit time-constant is 10 m sec for the values of fig. 15*a*. This makes the trigger circuit the limiting factor; the value of 10 m sec is too long for a practical radiation monitor. This difficulty was overcome by connecting C_t to the anode instead of to ground (fig. 15*b*). If $C_a \gg C_t$, then the anode waveform is little affected, while the trigger — after a depression — is restored through positive ion collection by the end of the anode swing to a potential normal for this stage of the cycle. Thereafter anode and trigger potentials rise together and the whole circuit is re-set.

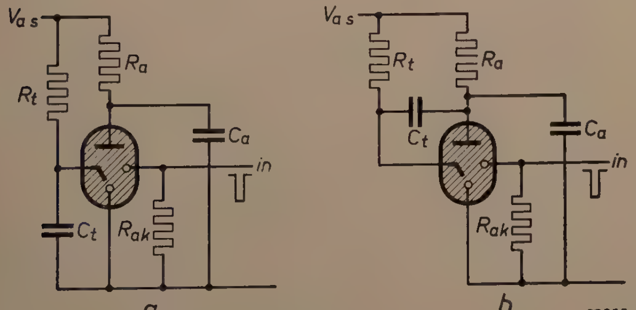


Fig. 15. *a*) Negative pulse rate meter circuit.
b) Negative pulse circuit with improved frequency response. Typical component values: R_t = 100 MΩ, C_t = 100 pF, R_a = 200 kΩ, C_a = 0.01 μF, R_{ak} = 1 MΩ, C_{ak} = 15 pF.

[11]) F.S. Goulding and C.H. Tosswill, British Patent No. 723 892.

Until the above circuit change was made it had been assumed that a return to the normal input sensitivity was dependent only upon completion of circuit recovery together with the re-establishment of the priming current. Moreover, it seemed likely that incomplete de-ionization would make the circuit exceptionally responsive for a period. These expectations, however, were not borne out by the results. *Fig. 16* shows the variation of the required Q_{in} with time elapsed since the last input pulse. Circuit equilibrium is restored after about 1 msec, but even after 10 msec sensitivity is still rising slightly.

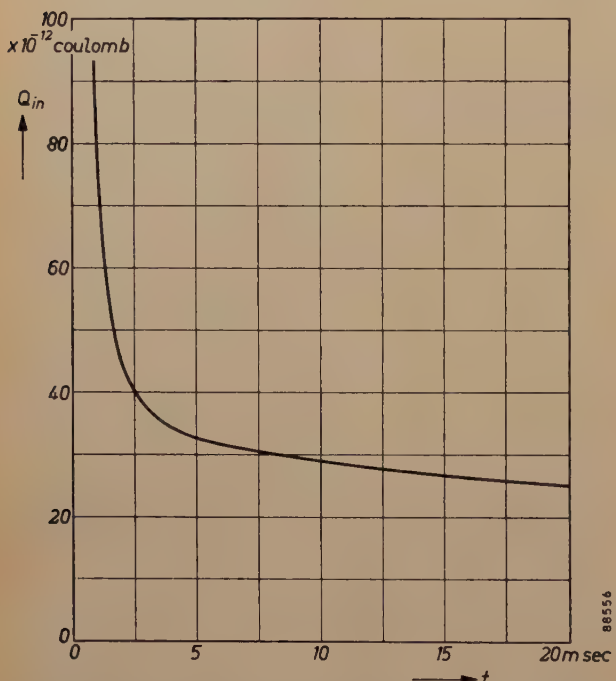


Fig. 16. Recovery of input charge sensitivity. About 1 millisec after the main discharge has passed, the charge sensitivity of the tube has reached a value of about 5×10^{-11}. At 10 millisec the sensitivity is 3×10^{-11} and still improving slightly.

The conclusion seems to be that the correct action of the Z 801U tube depends on the development of a very small unstable discharge between trigger and main cathode, and that this unstable discharge cannot occur until the general level of ionization in the tube has fallen too far for the trigger-resistor current I_{tb} to be provided by ion collection alone. De-ionization in this sense is likely to be protracted, being delayed by the persistence of metastable atoms and by a measure of ion replacement once circuit recovery is complete.

The cure for the delay is evidently to accelerate the process of de-ionization. This has been done in laboratory models having an electrode "labyrinth" in which the discharge passes through narrow passages rather than through a void. However, as the radiation monitor for which the Z 801U was developed can accept the frequency response provided by the standard structure (up to 1000 c/s), it has proved to be unnecessary to increase the frequency response up to 10 kc/s.

Charge amplification

In the rate meter circuit for which the Z 801U was developed, the function of the tube is to deliver a large fixed output charge every time the tube is switched by the arrival of a small input charge not less than 3×10^{-11} coulomb. So far as we are aware, this represents an input sensitivity far higher than that obtainable with any other type of cold-cathode tube. The tube provides, at low frequencies, an amplification factor of not less than 10^4, since Q_{out} may be as much as 5×10^{-7} coulomb.

If it is desired to make use of input pulses of one volt or less, then the effective input capacitance must exceed 30 pF. A capacitance of this size would upset the stability of the trigger circuit, and must therefore be isolated from the tube by means of selenium diodes [12]).

The radiation monitor which uses the Z 801U tube as a rate meter tube now also employs another in a variant of the Goulding stabilizer circuit to provide an output of about 155 volts.

The two tubes described above were developed specifically for their intended applications, viz. the stabilizer and rate meter circuits of a radiation monitor. A third tube, the Z803U, has, however, also been developed for more general application. In this tube the priming electrode is an anode, a priming discharge of some 10 μA flowing between it and the cathode in a region shielded from the trigger-cathode space. The tube has the same stable triggering characteristics as its predecessors; the nominal trigger breakdown voltage V_{tb} is 132 V. Anode breakdown voltage is 290 V, while the maintaining voltage V_{am} is 105 V. The tube can pass a mean current of 8 mA.

Another general purpose tube, type Z70U, of subminiature dimensions is in course of development at Eindhoven. This tube can be used with a higher supply voltage than the Z 803U but its mean current capacity is less. A further article will appear in due course in this Review describing these developments.

Acknowledgement should be made of the contribution to this work of Messrs E. Franklin,

[12]) See the article by F. S. Goulding referred to in [8]).

K. Kandiah and F. S. Goulding of the Electronics Division of A.E.R.E., Harwell, directed by Dr. D. Taylor.

Summary. After a brief introduction dealing with the operation and properties of cold-cathode trigger tubes, work is described on the development of two special trigger tubes for use in a radiation monitor. Two of the measures taken to improve reliability and performance are dealt with: the use of molybdenum sputtering to achieve better stability and the provision of a priming current to eliminate the statistical lag. Both tubes are designed to work in self-quenching circuits. The stabilizer tube, a tetrode with an auxiliary cathode, is designed for providing a stabilizing supply for Geiger tubes. The circuit conditions for satisfactory operation are described. The advantages of this stabilizer are its voltage range (200—300 V), its current range (5—150 μA) and the fact that the stabilized output voltage can readily be adjusted by changing the values of two resistors. The rate meter tube is designed to handle Geiger tube pulses. It is also a tetrode but differs from the stabilizer tube in that its priming electrode, again an auxiliary cathode, lies in the path of the main discharge and provides a "direct" priming discharge. The special characteristic of this tube is its extremely high sensitivity; in the negative pulse circuit, the input charge sensitivity is 3×10^{-11} coulomb. This tube can also be used for voltage stabilization.

AN AUTOMATIC NOISE FIGURE INDICATOR

621.317.34:621.396.822

The output signal from any amplifier is inevitably accompanied by a certain kind of disturbance in the same frequency range as the signal. Amongst the causes of this kind of disturbance are the thermal motion of the electrons in resistors and irregularities in the transit time of electrons in valves. Such disturbances are given the general name of *noise*, irrespective of their origin; their amplitudes have a random distribution, that is, in accordance with a Gaussian curve.

The extent to which an amplifier produces noise is expressed by the *noise factor* F. For a linear amplifier (having an impedance Z_i between its input terminals) F is defined, for a frequency f, as the ratio of the noise-output power ΔP_n in the frequency interval $f \pm \frac{1}{2}\Delta f$ to that part of ΔP_n caused by the thermal agitation noise of Z_i when Z_i is at a temperature of 290 °K (17 °C).

The noise factor is a function of frequency. To make it possible none the less to characterize the noise properties of an amplifier by a single quantity, the *mean noise factor* F_m has been introduced. This is by definition [1]) the ratio of the *total* output-noise power P_n to that part of P_n caused by the thermal agitation noise of Z_i, Z_i again being at a temperature of 290 °K. Whenever mention is made of the noise factor in what follows, it is the *mean* noise factor F_m that is meant.

It is often useful to be able to measure the noise factor of, for example, amplifiers in the relay stations of short-wave radio links, or that of equipment for radio-astronomy [2]). One method of measuring the noise factor is based on the fact that the power of the thermal noise produced by a resistor is proportional to the absolute temperature of the resistor. In this method the output noise power is measured (with a thermocouple) with two input resistors: first with a resistor that has a suitable resistance at room temperature, and then with a resistor that has the *same* resistance at another known temperature, e.g. that of liquid nitrogen. From the two temperatures, the resistance, and the two output noise powers, the noise factor can then be readily evaluated.

Another well-known method is based on the principle of passing a known noise current through the input resistor, e.g. the current flowing through a saturated diode [3]). The r.m.s. value of i, the fluctuations in the current, is determined by

$$\overline{i^2} = 2eI\Delta f, \quad \ldots \ldots \quad (1)$$

e representing the charge on the electron and I the mean value of the diode current. If we connect a resistor R_i, whose resistance is small compared to the internal resistance of the noise diode, between the input terminals of the amplifier, and then pass through it a noise current of such intensity that the noise power at the output is exactly doubled, the noise figure is given by:

$$F_\mathrm{m} = \frac{e}{2kT}\, I R_\mathrm{i},$$

where k is Boltzmann's constant and T the absolute temperature of the resistor R_i. Inserting the numerical values of the constants e and k ($e = 1.60 \times 10^{-19}$ coulomb, $k = 1.38 \times 10^{-23}$ joules/°K) and 290 °K

[1]) This definition is in accordance with American Standard 53 IRE 7 S1. Other definitions may be found in the literature; see for example Philips tech. Rev. 2, 329, 1937; 6, 133, 1941; 11, 84, 1949/50.

[2]) See for example C. A. Muller, Philips tech. Rev. 17, 305-315 and 351-361, 1955/56.

[3]) See for example Philips tech. Rev. 2, 330, 1937.

for T, we find:

$$F_m = 20.0\, I\, R_i \quad \ldots\ldots\ldots \quad (2)$$

(For frequencies above about 100 Mc/s a correction to formulae (1) and (2) is necessary in view of the transit time effect; absolute calibration with a resistor as noise source is then advisable.)

Instead of doubling the output noise power by the addition of diode noise, another possibility is the temporary reduction of the voltage amplification of the amplifier under test by a factor of $\sqrt{2}$, whereupon the output noise power is brought back to its initial value by the addition of diode noise. The power of the added noise is again a measure of the noise factor (provided of course that the variation of the amplification is effected at a point of the circuit where it cannot perceptibly influence the noise factor or the bandwidth).

This last method might be carried out as follows: the bias of the output valve of the amplifier (the output valve, because it contributes hardly anything to the output noise) is made to alternate periodically between the value at which the amplification is A and the value at which the amplification is $A/\sqrt{2}$; we shall call these intervals the odd and the even half-periods respectively. Simultaneously with this variation of the amplification, the anode voltage of the diode is modulated in such a way that there is no diode current in the odd half-periods and the saturation current I in the even half-periods. The diode current can be controlled by altering its filament current, for it possesses a directly heated tungsten cathode. The filament current is now so adjusted that the demodulated output voltage of the amplifier becomes as low as possible. Once this condition is reached, the added diode noise just compensates the $\sqrt{2}$-fold decrease in the amplification; in accordance with (2) the diode current I is now a measure of the noise factor sought.

A direct-reading noise factor meter could be built on this principle, but its operation would be a fairly complicated business, since in each case the filament current would have to be adjusted by manual control to a value such that the demodulated output voltage was at a minimum. If the task is to bring down an amplifier's noise figure to the minimum, every alteration to a network parameter would necessitate measuring the noise factor in the manner described above, and if three or more network parameters had to be altered, the job of finding the optimum setting might be a matter of several days. The whole procedure could be considerably speeded up if there was an instrument available that would automatically indicate the noise factor without requiring any manual adjustment, thus giving a direct indication of the effect of every alteration in the amplifier [4]).

This aim may be achieved by providing the above-mentioned arrangement with *automatic* regulation of the filament current to the value at which the demodulated output voltage is at a minimum [5]). *Fig. 1* shows curves for this voltage according as the diode current I is (a) too low, (b) at the right value and (c) too high; it is a square-wave voltage of 400 c/s (this is the frequency selected for varying the amplification and for modulation of the diode), which passes through zero and changes its phase when I passes the correct value. In order to obtain an effective system of control, the filament current for the noise-producing diode is likewise given a frequency of 400 c/s, and supplied by an auxiliary amplifier controlled by the demodulated output voltage. (For this a *selective* amplifier tuned to 400 c/s is necessary, since only the component of 400 c/s must be amplified.) If I is too small, then the filament current, and hence also I, are automatically increased; if I is too large, the filament current is reduced.

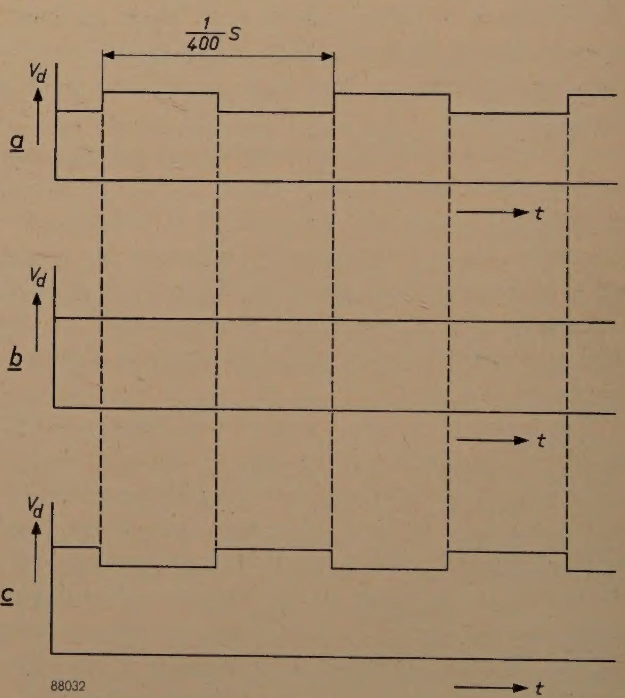


Fig. 1. Demodulated output voltage when the filament current of the noise-producing diode is (a) too low, (b) of the right value, and (c) too high. The frequency with which the amplification is varied and the noise diode is modulated is 400 c/s.

[4]) Our thanks are due to Mr. C. Seeger (Netherlands Foundation for Radio Astronomy, Leiden) for drawing our attention to the importance of such an instrument and for discussing its initial design with us.

[5]) Since the conclusion of this work two papers have been published (American patent 2 691 098 of October 5, 1954, in the name of W. Selove, and an article by H. Wallmann in Acta Polytechnica (El. Eng. Series) 6, 1955, No. 6), in which instruments for the same purpose are described, with minor differences of design.

As in most automatic control systems, the "actual value of the variable quantity" is compared with the "desired value", whilst the difference between the two serves for adjusting the "control element". In our case the actual value of the initial quantity is the output voltage during the even half-periods; the desired value is the output voltage during the odd half-periods; the difference is the block-shaped demodulated output voltage; and the control element is the noise-producing diode.

In principle any value may be selected for R_i, provided that it is small compared to the internal resistance of the noise-producing diode. 50 Ω is suitable, since it is a standardized value in all equipment for decimetric and centimetric waves in most parts of the world. With $R_i = 50$ Ω, equation (2) shows that the noise figure is just equal to twice the diode current in mA read on the meter (this current flows only during half the time, so that I is twice the meter reading).

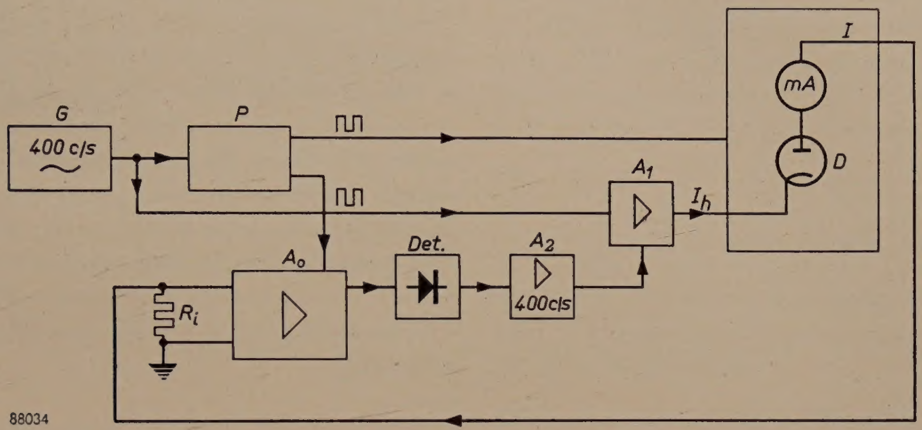


Fig. 2. Block diagram of the automatic noise factor indicator. The RC generator G produces a sinusoidal voltage of 400 c/s. A_1 amplifier producing the filament current I_h for the noise diode D. P pulse shaper supplying a square-wave voltage of 400 c/s. A_0 amplifier whose mean noise factor is to be measured. R_i input resistance. Det detector. A_2 selective amplifier, which passes only a narrow band at 400 c/s and controls the amplification of A_1. The mean noise factor is read on the milliammeter mA.

The working principle is shown in the block diagram of *fig. 2*. G is an RC generator, supplying a sinusoidal voltage of 400 c/s. The amplifier A_1 connected to it supplies the filament current I_h for the noise-producing diode D. Also connected to G is a pulse shaper P, in which the sinusoidal voltage is amplified and clipped, producing a square-wave voltage. The main function of this voltage is to give the periodic variation in the amplification of A_0, the amplifier under test, between the values A and $A/\sqrt{2}$; this is effected by applying the square-wave variable-amplitude voltage to the cathode of the last tube of A_0. The square-wave voltage also effects the synchronized opening and closing of the noise-producing diode. During the even half-periods the latter produces a noise current through the input resistance R_i of the amplifier A_0. The selective amplifier A_2 amplifies only the 400 c/s component of the output voltage demodulated by the detector *Det*, and regulates the amplifier A_1 in such a way that it produces approximately the correct filament current. The correct value is approached the more closely as the overall amplification in the closed loop A_0-Det-A_2-A_1-D-A_0 is greater.

Fig. 3 is a photograph of an instrument constructed on the principles described above.

Fig. 3. Automatic noise factor indicator constructed in accordance with the block diagram of fig. 2. (The instrument is not being manufactured.)

As regards the accuracy of the measurements given, this is in principle limited by the fact that the output voltage during the even half-periods can only be made to approximate to that during the odd half periods, since both are noise voltages. The error thus arising can be reduced by narrowing the bandwidth β of the selective amplifier A_2 (fig. 2). (Strictly speaking, it should be taken into account that the time constant of the filament of the diode is a contributary factor in determining the bandwidth after detection.) Thus, when a noise spectrum, assumed to be of rectangular shape (bandwidth Δ), is demodulated and the result is passed through a filter (bandwidth $\beta \ll \Delta$), we obtain a voltage whose r.m.s. value V shows the relative fluctuation [6])

$$\frac{\Delta V}{V} = \frac{\pi^{\frac{3}{2}}}{2\sqrt{2}}\sqrt{\frac{\beta}{\Delta}} \approx 2\sqrt{\frac{\beta}{\Delta}}.$$

In amplifiers for decimetric waves (for which our instrument has been primarily developed) Δ may be, say, 5×10^6 c/s; if $\beta = 5$ c/s, $\sqrt{\beta/\Delta}$ is of the order of 10^{-3}, so that the error does not exceed a few tenths percent.

For frequencies above 1000 Mc/s, the high-vacuum diode with tungsten filament is not the most suitable noise source. In such cases is it better to use a gas-discharge tube [7]), e.g. of the types K 50 A or K 51 A, which are specially developed for noise measurements. With these tubes the noise cannot be controlled by the filament current. The circuit described above, however, may be modified so that, instead of the filament current, the ratio of the duration of the odd to that of the even intervals is controlled. The even intervals (during which noise is added) should then be made so long that they correspond to a noise factor greater than that to be measured. Since the gas-discharge tube has to operate continuously, one must employ either an amplifying tube with an alternately amplifying and suppressing action (gating tube), or an attenuator with electrically varied attenuation. Further details concerning these two possibilities will not be dealt with here.

[6]) This formula follows from calculations by R. H. Dicke, Rev. sci. Instr. 17, 268-275, 1946.

[7]) K. S. Knol, Philips Res. Rep. 6, 289-302, 1951.

F. L. H. M. STUMPERS and N. van HURCK.